# AN INTRODUCTION TO EDUCATIONAL STATISTICS

C. W. ODELL
COLLEGE OF EDUCATION
UNIVERSITY OF ILLINOIS

New York PRENTICE-HALL, INC.

First Printing.........January, 1946
Second Printing.......January, 1951

PRINTED IN THE UNITED STATES OF AMERICA

# Preface

WHEN the writer prepared *Statistical Method in Education*, he had in mind the needs of those institutions that offer more than one semester's work in the subject and of those individuals who wish to prolong their study beyond a first course. Experience has revealed, however, that a more limited treatment finds favor with many persons, since they do not desire to pursue the subject beyond what can be covered in a one-semester course and therefore do not care for a larger and more comprehensive book than is necessary for that purpose. He has written this volume for such persons, including in it only as much as his students cover in one semester. He has added a few suggestions for further study, so that those who wish to carry on will know where to turn for assistance.

Much of the content is relatively well known; new formulas, procedures, and other matters appearing appropriate in an elementary text have been added. The presentation avoids higher mathematics, presupposing no preparation therein beyond arithmetic and a small amount of high-school algebra and geometry. The chief abilities needed to master the content are the power to think clearly, the accurate and reasonably rapid performance of the fundamental operations with integral and decimal numbers, and the habit of careful, orderly work.

The content is arranged in the order which the writer prefers, but instructors may wish to make changes. For example, Chapters XIII and XIV contain material part or all of which may be introduced earlier with little difficulty. For those who desire to cover even less material than the volume presents, the writer suggests Chapters XI and XII and the topics on the geometric mean, the harmonic mean, estimated true scores, determining difficulties of test elements, and equally noticed differences as portions that may be omitted.

C. W. ODELL

# Contents

# List of Figures

# List of Tables

# CHAPTER I

# Introduction

## Purpose and plan of this treatise

In preparing this volume the author has had in mind that there are three chief groups of persons engaged in educational activities for whom acquaintance with statistical methods is essential. One is composed of research workers, that is, producers. A very necessary step in the development of any science is the application of methods of measurement and computational procedures to experimental and other research work therein. Recognition of this fact, with consequent development and application of statistical methods, has come rather recently in education and the other social sciences, but it is now widespread. No one can be considered competent to conduct many types of research unless he has command of statistical principles and procedures. Their use renders it possible to give brief, exact descriptions and convenient, meaningful summaries; tends to make thinking definite; provides bases for drawing conclusions and for making estimates; and enables workers to analyze and simplify complex data.

A second and much larger group is that of teachers, administrators, and others who devote their efforts chiefly to the day-by-day operation of the educational system. These consumers need both the ability to select and employ those statistical methods which contribute to the efficient performance of their duties and also sufficient understanding of statistical terminology and procedures so that they can mentally digest and then apply the results of studies carried on by others. Modern educational literature is replete with statistical terms. Such terms, as well as the applications of statistical procedures, occur

in almost all aspects of educational activity. They are found in studies and discussions of the curriculum and the extracurriculum, of buildings and equipment, of pupils and teachers, of textbooks and other supplies, of finances, and so on almost without end. They are frequent in school surveys, in guidance and counseling, in diagnosis and prognosis, in reports on absence and tardiness, in studies of progress and elimination, and so forth. Indeed, no important phase of our school system operates without some application of statistics.

The third group is composed of students, chiefly graduate but also undergraduate. Their needs are similar to those of the second and also, for many individuals, include those of the first group. It is for this last category, including both present and future consumers and producers, that this volume is chiefly intended. It is, therefore, its purpose to present as many fundamental procedures, principles, and terms as the typical class of advanced undergraduate or graduate students without especial training in mathematics can cover in a one-semester course. The treatment has been made as nonmathematical as is consistent with the objectives and content, emphasizing practical computation, use, and understanding, rather than derivations and theoretical considerations. The mathematical ability most needed is that to perform the four fundamental operations accurately and fairly rapidly with integral and decimal numbers. Also, a few elements from high-school algebra and plane geometry are necessary, but no mathematics beyond that.

The author hopes that all those who study this treatise will gain therefrom some enhancement of each of the several abilities listed below:

To select appropriate procedures and measures;
To compute measures;
To interpret measures reported by others;
To associate terms and concepts;
To present results.

## General suggestions concerning statistical procedure

Statistical method is not a substitute for thought, but an aid—often an essential aid—to thought. It is a means, not an

end. It should be employed only to the extent needed for the solution of the problem at hand or for the conclusion of the desired research. Graduate students and other beginners in educational research are frequently prone to employ unnecessarily elaborate and complicated statistical procedures in the belief that so doing increases the value of their results. Instead, such statistical measures and computations as most directly and simply reveal the significant facts should be used.

One violation of the principle just stated is the attempt to correct the inadequacy of data by overrefined statistical procedures. In general, measures derived from data are no more reliable and valid than the original data. Certain exceptions to this will be noted in appropriate connections later. They are chiefly concerned with types of errors the effects of which tend to balance one another or to be negligible when some measures are computed. Moreover, increasing the number of cases may or may not serve to eliminate errors, depending upon what type of errors is involved.

In general, the statistical work involved in a study should be planned before it is begun. The plan should provide for collection of data, preliminary analysis and tabulation thereof, classification and subsequent tabulation, computation, interpretation, and presentation. Minor modifications will often prove desirable; but, if a worker finds that he is constantly making major additions or alterations to his originally planned statistical procedures, he may be sure that his initial planning has been too brief or otherwise unsatisfactory. He should consider carefully such matters as the kind of data to be used; how they are or may be interrelated; how they bear upon the question or questions at issue; how accurate, adequate, representative, and otherwise valid they are; how they are to be analyzed; what measures can and should be computed from them; how they can be so collected, expressed, and tabulated as to be most conveniently employed; to whom the conclusions are to be presented and therefore in what tabular, graphic, and other forms they should be arranged; and other similar points.

No one can become a competent statistical worker who does not possess the capacity, both intellectual and temperamental,

for careful, thorough, unbiased, accurate work. This includes both the computational and the reasoning aspects of the work. An individual competent in only one of these two phases cannot become an efficient statistician. His best recourse is to collaborate with someone competent in the other phase—an arrangement that may be satisfactory. Moreover a worker cannot employ statistical method constructively unless he has a wide knowledge of the field and can think imaginatively. Insofar as the degree of computational ability necessary to accurate use of the formulas and methods given in this book is concerned, almost every student at all likely to be studying educational statistics can develop it by sufficient practice.

## Computation

Even though the ability to compute accurately and with fair rapidity is essential for efficient statistical work, anyone who expects to do much in this field should expedite his labors by employing computational aids. These fall into two general classes: tables and mechanical devices. Numerous types of both are available.

The one most useful set of tables for workers in educational statistics is probably that known as Barlow's.[1] This gives the squares, cubes, corresponding roots, and reciprocals of all integers up to 12,500. By interpolation, both nonintegral and larger numbers may be similarly dealt with. Less extensive tables may be almost as useful in elementary work. Multiplication tables extending up to such limits as 1,000 × 1,000 or 100 × 10,000 are available, but have not received much use among American statisticians. In some computations, not many of which are among those included in this book, logarithmic tables are helpful; in a few they are indispensable.

In addition to tables such as have been mentioned, which are widely useful in numerous types of work, there are others of more limited scope insofar as their area of usefulness is con-

[1] Barlow, Peter, *Tables of Squares, Cubes, Square Roots, Cube Roots, and Reciprocals of All Integer Numbers Up to 12,500.* Fourth Edition. London: E. and F. N. Spon, 1941. 258 pp.

cerned. Pearson's[2] sets of tables include a number concerned with functions of the normal curve and also others frequently employed. One of them—Sheppard's Table of the Probability Integral—is the commonly used source table for many procedures and computations.

Still more limited are groups of tables prepared specifically for the use of educational, or educational and psychological, workers. These groups tabulate values for formulas, operations, and functions frequently needed by such persons. Many statistical texts contain one or more tables of this sort. Three sets will be mentioned here. Dunlap and Kurtz[3] are the authors of a handbook of three parts that deserves high rank. The first part presents almost thirty nomographs, or graphic computational aids, useful for securing values often desired in educational work. The second contains a dozen tables, some of general computational nature and others more or less peculiar to education. The third part lists more than four hundred formulas and two hundred symbols used in this field. Holzinger[4] has compiled a set of a dozen tables, similar to those in the second part of Dunlap and Kurtz's book, that represent a well-chosen combination of the two types. Kelley's[5] set has only five tables; but, since the most important of them—Number I—contains values of several functions, they cover about as wide a field of usefulness as those of Holzinger.

The chief mechanical aids to computation are slide rules and computing machines. For those familiar with their use, slide rules, such as engineers commonly employ, are quite helpful in multiplication, division, and some other operations. A ten-inch rule gives approximate accuracy to three figures; a twenty-inch rule, to four or five figures. Dunlap and Kurtz have devised an inexpensive slide rule which may be used as

---

[2] Pearson, Karl, Editor, *Tables for Statisticians and Biometricians.* Cambridge: Cambridge University Press. Issued by Biometric Laboratory, University College, London. Part I, Second Edition, 1924. 143 pp. Part II, 1931. 262 pp.

[3] Dunlap, Jack W., and Kurtz, Albert K., *Handbook of Statistical Nomographs, Tables, and Formulas.* New York: World Book Company, 1932. 163 pp.

[4] Holzinger, Karl J., *Statistical Tables for Students in Education and Psychology.* Fifth Impression. Chicago: University of Chicago Press, 1938. 104 pp.

[5] Kelley, Truman Lee, *The Kelley Statistical Tables.* New York: The Macmillan Company, 1938. 136 pp.

suggested above and also for finding values of a number of functions often needed in educational and psychological work.[6] Enlow has constructed a more elaborate and durable—also more expensive—rule of the same general type.[7]

Some computing machines are hand-operated; others are electrically driven. Those of the latter type are superior but much more expensive. The machines vary in scope of usefulness from simple adding machines to those with such features as automatic multiplication and division. Among the leading varieties are the Comptometer, the Burroughs, the Dalton, the Friden, the Marchant, and the Monroe.

Still more complicated than these, and more useful for certain operations, are the card-punching and automatic sorting machines, such as the Hollerith and that of Warren and Mendenhall. For a project involving multiple classification of many cases such a machine is highly desirable. It will not only sort the data, but also will yield results that greatly facilitate computing coefficients of correlation and other frequently wanted measures.

A still different variety of machine is that for scoring tests. Several of this sort have been devised, but the only one receiving very much use is that of the International Business Machines Corporation. This electric machine will score special answer sheets designed for use with numerous standard tests with high accuracy and at a rate of several hundred per hour.

## Accuracy

The accuracy of results obtained by statistical procedures depends upon the accuracy of both the original data and the computations to which they are subjected. Before actually performing the statistical operations involved in any project, the computer should decide how accurate he wishes the results to be and then take such precautions in his work as will, if possible, insure that degree of accuracy but at the same time avoid such unnecessary labor as carrying more significant figures or decimal places than are essential.

---

[6] This rule may be obtained from the Psychological Corporation.
[7] Enlow's slide rule may be secured from Keuffel and Esser.

In the cases of some types of data, common usage has more or less agreed upon the degree of accuracy to which they are determined and reported, but in most instances this decision rests with the person making the study. The person or persons for whom the conclusions are intended is often an important factor. Data designed for the general public or other large lay groups should usually be "rounded off," so as to make them easy to comprehend and remember; whereas those for small, highly interested groups of laymen or for professional workers frequently may well be more nearly exact. Too great detail may confuse rather than inform the lay or relatively uninterested hearer or reader. This is particularly true of numerical oral statements. If an audience is told that the average salary of teachers in one system is about \$1,250, for example, whereas that in a group of comparable systems is about \$1,400, they will probably get and retain a clearer impression than if they are given such figures as \$1,245.18 and \$1,406.66, respectively. Similarly, if they are told that enrollment is increasing about 5 per cent per year, they will be likely to receive and remember the point better than if they are told that the annual increase is 4.86 per cent.

There are numerous rules as to the steps necessary to insure desired degrees of accuracy in results and, conversely, as to how great accuracy results from specified procedures. At the risk of overemphasizing the point, a few of them will be given here. They may be stated in terms of either *absolute* or *relative* accuracy. Sometimes one is more convenient, sometimes the other. Absolute accuracy is expressed in terms of the actual amount; relative accuracy, in terms of what fraction that is of the number itself. Several examples of absolute accuracy expressed decimally will be given here and some of relative accuracy later. Thus 4.6 is conventionally understood to be accurate to one decimal place or to the nearest tenth, that is, to represent a value that, if more accurately given, would be nearer to 4.6 than to either 4.5 or 4.7. Another way of stating the same thing is that 4.6 represents a true value between 4.55 and 4.65. Therefore its limit of error is .05. Similarly 23.08 is accurate to two decimal places, or to the nearest hundredth,

or represents a true value between 23.075 and 23.085; 573. is accurate to the nearest unit, or represents a true value between 572.5 and 573.5; and so on.

If a whole number ends in one or more zeros and is accurate to the nearest unit, it should be followed by a decimal point; if not, it should be understood as accurate only to the last digit, not zero. Thus 2,600. should be understood as accurate to the nearest unit, that is, as representing a true value between 2,599.5 and 2,600.5; whereas 2,600 should be considered accurate only to the nearest hundred, or as standing for some value between 2,550 and 2,650. Sometimes the notation $26. \times 10^2$ is used to mean the same thing. When such a number ends in two or more zeros of which some are accurate and some not, the latter method may be employed, or a dot or bar may be written over the last significant one. Thus $260. \times 10$, or $2{,}6\dot{0}0$ or $2{,}6\bar{0}0$, denotes accuracy only to the nearest ten, that is, a true value between 2,595 and 2,605. Similarly $520. \times 10^3$, or $52\dot{0}{,}000$ or $52\bar{0}{,}000$, denotes accuracy to the nearest thousand, or a true value between 519,500 and 520,500.

Absolute accuracy may also be expressed in terms of *significant digits* or *figures*. A number has as many significant figures as it contains digits that are considered correct. Zeros used merely to indicate the position of the decimal point are not counted. Thus 234. and 23.4 and .00234 each has three significant figures; 25,000 and 6.8 and .0090 each has two; 47,914 and $12{,}58\dot{0}{,}000$ and .074208 each has five; and so on with others.

The relative accuracy of a number is the limit of accuracy or error divided by the number itself. Thus, since 4.6 represents a true value between 4.55 and 4.65, its error cannot be greater than .05 and its relative accuracy is $.05 \div 4.6$, which gives .0109 or 1.09 per cent. Since 23.08 denotes a value between 23.075 and 23.085, its limit of error is .005 and its relative accuracy is $.005 \div 23.08$, which equals .000217 or .0217 per cent.

The error in a sum or total is equal to the algebraic sum of the errors in the quantities added to produce it. In other words, it equals the difference between the sum of the posi-

tive errors and that of the negative errors. If all the errors are in the same direction,[8] their algebraic sum is the same as their arithmetical sum; hence the error in the total equals the latter. If, as often, the errors in the addends are as likely to be positive as negative,[9] their total tends to approximate zero; so the sum is likely to be rather accurate. In general, a total is more accurate than most of the quantities that contribute to it. Its last certainly accurate figure, however, is no farther to the right than the last one in the least accurate of the measures added. For example, the sum of the numbers at the right cannot be depended upon beyond 132.6 and should be given in that form, not as 132.608. The same is true of a remainder or difference.

| |
|---:|
| 41.875 |
| 13.2 |
| 8.34 |
| 23.213 |
| 9.5 |
| 36.48 |
| 132.608 |

Since an average is a sum divided by the number of items added, its absolute error is the error of the sum divided by the number of cases. Hence, if the errors in the items are variable, the error in the average is likely to approach zero. This fact justifies carrying an average to more decimal places than are given in the items from which it is secured. If fewer than 10 items are added, their average should not be carried to more decimal places; if from 10 to 99, it may be carried to one more; if from 100 to 999, to two more; if from 1,000 to 9,999, to three more; and so on. In other words, the additional decimal places to which an average may be carried are one less than the number of digits in the number of cases. Thus, since there are three digits in 450, an average of 450 cases may be carried to two more places than the single items; since there are two digits in 38, an average of so many may be carried to one more; and similarly for others. Another rule that is practically the same in effect is that generally as many figures are accurate in an average as in the sum from which it was computed.

The most commonly given rule for multiplication and division is that the number of significant figures in a product or

---

[8] Such errors are called *systematic, constant, group, biased, cumulative,* or *noncompensating.*

[9] Errors of this sort are known as *variable, chance, accidental, individual, random, unbiased, noncumulative,* or *compensating.*

*quotient is generally the same as in the one of the two contributing numbers that has the fewest. Frequently the last* may be in error by 1, however, and occasionally by more. For example, the product of a multiplicand with four significant figures and a multiplier with two generally has two accurate; the quotient of a dividend with three and a divisor with five ordinarily has three; and so on. As an illustration, 24.71 has four significant figures accurate and 1.8 has two; hence their apparent product, 44.478, is probably accurate only to two places—44. This can be tested by multiplying together the minimum and also the maximum values that the numbers may represent. The former, 24.705 and 1.75, give 43.23375, and the latter, 24.715 and 1.85, give 45.72275. Both differ from 44. in the last figure. As another example, $75.94 \div 800. = .094925$, which is probably accurate as far as .0949. To test this, as in any case of division, the minimum value of the dividend should be divided by the maximum value of the divisor, and vice versa. This gives $75.935 \div 800.5 = .094859$ and $75.945 \div 799.5 = .094991$. The first of these reduced to four places agrees with .0949, whereas the latter becomes .0950. Since a power of a number is obtained by multiplication, this same general rule applies to it also.

The accuracy of products and quotients may be dealt with relatively as well as absolutely. The limit of error in a product or quotient is the approximate algebraic sum of the limits of errors in the numbers involved, all errors being relative. For example, $1.5 \times 40. \times .0648 = 3.888$. The limits of relative errors of the three numbers are $\frac{.05}{1.5} = .0333$, $\frac{.5}{40.} = .0125$, and $\frac{.00005}{.0648} = .0008$, respectively. Their sum is .0466; hence this is the approximate limit of the relative error of the product, 3.888. If it is desired, the approximate limit of its absolute error can be found by multiplying .0466 by 3.888, which gives .1812.

Powers and roots are special types of multiplication and division, and so may be treated as suggested above. Simpler procedures are available, however. A power usually has as

many figures accurate as its base and a root as many as the number from which it was extracted. The relative error of the $n$th power of any number is approximately $n$ times that of the number itself, and the relative error of the $n$th root of any number is approximately $\frac{1}{n}$ that of the number. For example, if a number has a limit of error of 1 per cent, its square has a limit of error of approximately 2 per cent and its square root of approximately .5 per cent; if it has a limit of error of .3 per cent, its cube has an approximate limit of error of .9 per cent and its cube root of .1 per cent.

With logarithms, there is no gain in accuracy in using mantissas containing more than one more decimal place than the number of significant figures accurate in the number with the fewest. Thus, if the number with the least has three significant figures accurate, four-place logarithms are sufficient.

In a computation involving several steps, it is usually best to retain more figures or decimal places than are desired accurate in the final result, but application of the rules suggested and experience should guide workers in not burdening themselves by keeping more than are needed. When digits are dropped, the last preceding digit should be increased by one if the first of those dropped is 5 or greater. For example, if .4174 is reduced to two decimal places, it becomes .42; if .03846 is reduced to three, it becomes .038; if 24,252 is reduced to three significant figures, it becomes 24,300 or $243 \times 10^2$; and so on. In determining per cents and other quotients, the process of division should always be carried far enough to indicate whether the last significant figure or decimal place to be kept is correct as first found or should be increased by 1. Thus, if one wishes $10 \div 60$ to the nearest hundredth, he should not stop with .16, the first two figures, but should go on to .166, which indicates that the result should be written as .17, not .16.

We should not forget that numbers are often exact, in other words, accurate to an infinite number of places. For example, if one room has 41 pupils, another 38, and another 34, these numbers are exact, not approximate. Their sum is exactly 113; hence their average may be carried to as many places as desired.

A multiplier or divisor is frequently perfectly accurate, in which case accuracy of the multiplicand or dividend determines that of the product or quotient.

## References for further study

Although this volume deals with statistics not requiring much mathematical ability above the elementary-school level and none above that of the high-school level, yet some otherwise competent individuals have difficulty with that. Those who know or fear that they will have such a handicap are advised to secure and study Walker's[10] excellent book, which contains a series of short chapters on the mathematics essential to satisfactory work in elementary statistics. Each chapter begins with a self-test which students are to take. A high score thereon indicates that a student does not need to study the chapter in question; a low score, that he does. In the latter case, after study, he takes a test found at the end of the chapter to determine if he has mastered it sufficiently well and, if not, studies it further. The topics with which the chapters deal are such as significant figures, short cuts in computation, common fractions, multiplication of polynomials, and logarithms.

Kurtz and Edgerton's[11] dictionary of terms employed in the general field of statistics—mathematical, business, educational, psychological, and so forth—should be useful to many who use this volume. It contains definitions and explanations of over two thousand terms, symbols, and abbreviations. Another source in which may be found a number of definitions and critical discussions of the value and use of statistical procedures is the *Encyclopedia of Educational Research*, edited by Monroe.[12] Still another is the *Dictionary of Education*, compiled under the editorial leadership of Good.[13]

---

[10] Walker, Helen M., *Mathematics Essential for Elementary Statistics.* New York: Henry Holt & Company, Inc., 1934. 246 pp.

[11] Kurtz, Albert K., and Edgerton, Harold A., *Statistical Dictionary of Terms and Symbols.* New York: John Wiley & Sons, Inc., 1939. 191 pp.

[12] Monroe, Walter S., Editor, *Encyclopedia of Educational Research.* New York: The Macmillan Company, 1941. 1344 pp.

[13] Good, Carter V., Editor, *Dictionary of Education.* New York: McGraw-Hill Book Company, Inc., 1945. 495 pp.

There are numerous other volumes that those who wish to consult other books covering more or less the same range of content as this may use. Of these, the following are among those which the present writer regards as most helpful:

Broom, M. E., *Educational Statistics for Beginning Students.* New York: American Book Company, 1936. 318 pp.
Cooke, Dennis H., *Minimum Essentials of Statistics.* New York: The Macmillan Company, 1936. 271 pp.
Garrett, Henry E., *Statistics in Psychology and Education.* Second Edition. New York: Longmans, Green & Company, Inc., 1937. 491 pp. (This also contains more advanced material.)
Gray, Clarence T., and Votaw, David F., *Statistics Applied to Education and Psychology.* New York: The Ronald Press Company, 1939. 278 pp.
Lindquist, E. F., *A First Course in Statistics.* Revised Edition. Boston: Houghton Mifflin Company, 1942. 242 pp.
Tiegs, Ernest W., and Crawford, Claude C., *Statistics for Teachers.* Boston: Houghton Mifflin Company, 1930. 212 pp.
Walker, Helen M., *Elementary Statistical Methods.* New York: Henry Holt & Company, Inc., 1943. 368 pp.

Those who wish to pursue the study of statistics beyond the elementary stage which this volume presents likewise have the possibility of choice among many books, monographs, articles, and other printed sources, in both educational statistics and general mathematical statistics. Among volumes of the former type which present material distinctly in advance of that in this book are:

Guilford, J. P., *Fundamental Statistics in Psychology and Education.* New York: McGraw-Hill Book Company, Inc., 1942. 327 pp.
Holzinger, Karl J., *Statistical Methods for Students in Education.* Boston: Ginn & Company, 1928. 372 pp.
Lindquist, E. F., *Statistical Analysis in Educational Research.* Boston: Houghton Mifflin Company, 1940. 266 pp.
Monroe, Walter S., and Engelhart, Max D., *The Scientific Study of Educational Problems.* New York: The Macmillan Company, 1926. 504 pp.

The following are more general or mathematical treatments of statistics. The book by Mode and Part I of Kenney's book

are relatively elementary; the others include much advanced material:

Ezekiel, Mordecai, *Methods of Correlation Analysis.* Second Edition. New York: John Wiley & Sons, Inc., 1941. 531 pp.

Fisher, R. A., *Statistical Methods for Research Workers.* Eighth Edition. Edinburgh: Oliver and Boyd, 1941. 344 pp.

Kelley, Truman L., *Statistical Method.* New York: The Macmillan Company, 1928. 390 pp.

Kenney, John F., *Mathematics of Statistics.* New York: D. Van Nostrand Company, Inc., 1939. Part I, 248 pp.; Part II, 202 pp.

Mode, Elmer B., *The Elements of Statistics.* New York: Prentice-Hall, Inc., 1941. 378 pp.

Peters, Charles C., and Van Voorhis, Walter R., *Statistical Procedures and Their Mathematical Bases.* New York: McGraw-Hill Book Company, Inc., 1940. 516 pp.

Wilks, S. S., *Mathematical Statistics.* Princeton, New Jersey: Princeton University Press, 1943. 284 pp.

A limited number of specific references will be given at the end of each chapter. They are chosen to serve both those who wish to read other discussions of the same scope as those in the text and also those who desire more advanced treatments of the same topics.

## EXERCISES AND PROBLEMS

**1.** Between what limits does the true value of each of the following lie? (*a*) 29.64; (*b*) .088; (*c*) 7,400; (*d*) 1.42; (*e*) 2,100; (*f*) 980.; (*g*) 4,216.; (*h*) .00067; (*i*) 400,000; (*j*) 3.7914.

**2.** How many significant figures are there in each number in Exercise 1?

**3.** What is the limit of the relative error in each number in Exercise 1? Carry the answers to two significant figures.

**4.** Which of the following would probably cause systematic errors and which variable errors? (*a*) Use of a yardstick an inch too long; (*b*) allowing too little time for a standard test; (*c*) employing various teachers' estimates of pupils' weights; (*d*) assignment of marks by a teacher who is a severe marker; (*e*) measurement of short periods of time by use of a watch without a second hand; (*f*) determination of average class scores by averaging scores of five pupils chosen at random from each class.

**5.** To how many more decimal places than the single cases may

the average of each of the following numbers of cases safely be carried? (*a*) 150; (*b*) 26; (*c*) 4,600; (*d*) 84; (*e*) 7,298.

**6.** How many significant figures are probably accurate in each of the indicated products and quotients? (*a*) $5{,}000 \div 21.5$; (*b*) $220 \times 14.78$; (*c*) $1.65 \times 82$; (*d*) $31.648 \div 5.69$; (*e*) $.0465 \div 70.$; (*f*) $1.20 \times 6.1$.

**7.** Write the product or quotient of each part of Exercise 6 only so far as it is probably accurate.

**8.** Give, to four decimal places, the approximate limit of relative error of the product or quotient of each part of Exercise 6.

**9.** What is the approximate limit of relative error of each indicated power or root, if each number has the limit of absolute error indicated by its form? (*a*) $\sqrt{38.28}$; (*b*) $\sqrt[3]{4{,}000}$; (*c*) $(.045)^2$; (*d*) $\sqrt[4]{.4862}$; (*e*) $(21)^3$; (*f*) $(68.0)^2$.

**10.** Write each of the following to only two decimal places: (*a*) .0971; (*b*) 25.963; (*c*) .6464; (*d*) 1.53497; (*e*) 418.305; (*f*) .6667.

## REFERENCES

Fisher, R. A., *Statistical Methods for Research Workers.* Eighth Edition. Chap. I. Edinburgh: Oliver and Boyd, 1941.

Garrett, Henry E., *Statistics in Psychology and Education.* Second Edition. Pages 10–13. New York: Longmans, Green & Company, Inc., 1937.

Guilford, J. P., *Fundamental Statistics in Psychology and Education.* Pages 7–12. New York: McGraw-Hill Book Company, Inc., 1942.

Kenney, John F., *Mathematics of Statistics.* Part I, Introduction. New York: D. Van Nostrand Company, Inc., 1939.

Lindquist, E. F., *A First Course in Statistics.* Revised Edition. Pages 1–6. Boston: Houghton Mifflin Company, 1942.

Mode, Elmer B., *The Elements of Statistics.* Chaps. I and II. New York: Prentice-Hall, Inc., 1941.

Monroe, Walter S., and Engelhart, Max D., *The Scientific Study of Educational Problems.* Pages 61–65. New York: The Macmillan Company, 1926.

Peters, Charles C., and Van Voorhis, Walter R., *Statistical Procedures and Their Mathematical Bases.* Pages 40–41. New York: McGraw-Hill Book Company, Inc., 1940.

Walker, Helen M., *Elementary Statistical Methods.* Chaps. I, VI; pages 11–23. New York: Henry Holt & Company, Inc., 1943.

# CHAPTER II

# Grouping Data

## Attributes and variables

Data are frequently thought of as being either attributes or variables, but it is better to think of dealing with them by the *method of attributes* or by the *method of variables.* The former involves *nonquantitative classification;* the latter, *quantitative.* Some data are of such nature that no quantitative or numerical basis of grouping them is valid; hence they must be handled by the method of attributes. Most of the data concerned in educational statistics are of such nature that they can be, and usually are, handled by the method of variables, but it is sometimes convenient to employ the method of attributes with them.

The method of attributes, also called *unordered, verbal,* or *qualitative* classification, frequently involves grouping data into only two classes, but may use more. Common examples where only two classes are used are the classification of pupils as boys or girls, of teachers as men or women, of curricula as college preparatory or noncollege preparatory, and of activities as curricular or extracurricular. Examples involving more than two-fold classification are of color of hair as black, dark-brown, light-brown, or red; of curricular as general, commercial, agricultural, technical, or classical; and of major subjects as history, English, Latin, mathematics, home economics, science, and so on. In some of these instances numerical classes might be employed, but generally they would be of doubtful validity and not in accord with common practice.

Among the data regularly dealt with as variables, or *ordered,* are school marks, test scores, heights, weights, ages, salaries,

enrollments, and many others. As suggested, any of these can be put in the form of attributes, but usually their quantitative nature is still evident. Thus marks may be classified as passing or failing; pupils as being overweight, at weight, or underweight; and so on.

When the method of variables is used, the number of groups or classes employed may range from very few up to a large number. In some instances the data are such that there cannot be more than a certain number of classes, but usually the number is determined by the judgment of the worker. The several chief considerations in such determination will be stated later in this chapter.

Variables may be either continuous or discontinuous. A *continuous variable* is one capable of infinitesimal subdivision. There are no impossible values between the highest and lowest points on the scale of measurement. For example, height is continuous. There is no linear distance so small but that the difference in height between two individuals or between two measurements of the same individual may be still less. The only limitation in this respect is in our ability to measure it. Frequently it is neither practicable nor worth while to measure variables nearly so exactly as we can. One-five-millionth of an inch has been measured, but in determining the heights of children accuracy to the nearest inch is usually sufficient. It is especially true in the cases of such traits as mental age, quality of handwriting, and others for which we do not have very accurate means of measurement that it would be useless, even ridiculous, to attempt very exact determination. Such a datum as a mental age of 10 years, 6 months, 16 days or a handwriting score of 74.35 per cent is, in the present state of educational measurement, an absurdity.

A *discontinuous* or *discrete variable*, on the other hand, is one that cannot be subdivided infinitesimally. Ordinarily the smallest possible difference in numbers of pupils in different rooms is one. If there are more than 33 pupils in a class, there must be at least 34.[1] In many instances there are artificial

[1] In exceptional instances, such as when a pupil has part of his work in one class, part in another, fractional numbers of pupils may be legitimate.

conditions which cause variables that are intrinsically continuous to be employed as discrete. Thus in spelling no finer unit than the word is commonly recognized, yet pupils' abilities may differ by only part of a word. If a system adopts a salary schedule with $50 increments, that amount becomes the least actual difference among salaries, although smaller ones are possible.

It is very general practice to treat discontinuous data as if they were continuous. Usually so doing introduces some error; but, if the least unit of difference practically possible is small in comparison with the data themselves, the error is likely to be negligible.

## Grouping or classifying data

All educational data, similar to those in other fields, are fundamentally collections of single scores, cases, or other items. If the number is quite small, there is rarely need for grouping them. They may be dealt with as individual items so easily that it is scarcely worth while to group them. Moreover, it is probable that grouping would result in appreciable sacrifice of accuracy.

Just how large the number of cases must be to justify grouping cannot be stated exactly. Whenever a worker cannot easily keep in mind and employ in his computations the individual items and can group them without serious loss of accuracy, he should do so. Thus a rural or small-town administrator with five teachers in his system can easily deal with their salaries by listing them, preferably in order. If he moves to a system with twice as many teachers, he can still do so. If he moves on to one with twenty teachers, the task becomes so difficult that he may need to group them, and if he advances to one with forty or fifty teachers, he is exceptional if he can deal with the total salary situation readily without grouping them.

To illustrate the fact just stated, the following salaries of the eighteen teachers in a certain high school may be used[2]: 1,400,

[2] The dollar mark has been omitted here and elsewhere where its frequent repetition would be cumbersome.

1,250, 1,900, 1,600, 1,550, 1,750, 1,650, 1,350, 1,800, 1,600, 1,450, 1,500, 1,700, 1,300, 1,450, 1,800, 1,600, 1,550. One can inspect these and easily determine the highest and lowest, but it is somewhat difficult to get and keep in mind an adequate conception of the whole group. If they are arranged in order, thus: 1,900, 1,800, 1,800, 1,750, 1,700, 1,650, 1,600, 1,600, 1,600, 1,550, 1,550, 1,500, 1,450, 1,450, 1,400, 1,350, 1,300, 1,250, the task becomes easier. If, however, the salaries of the thirty elementary teachers in the same system are added, the task becomes difficult, even if all are arranged in order of size. The entire forty-eight are: 1,900, 1,800, 1,800, 1,750, 1,700, 1,650, 1,600, 1,600, 1,600, 1,550, 1,550, 1,500, 1,500, 1,500, 1,500, 1,500, 1,450, 1,450, 1,450, 1,450, 1,450, 1,400, 1,400, 1,400, 1,400, 1,350, 1,350, 1,350, 1,350, 1,350, 1,350, 1,300, 1,300, 1,300, 1,250, 1,250, 1,250, 1,200, 1,200, 1,150, 1,150, 1,100, 1,100, 1,050, 1,050, 1,000, 1,000, 1,000.

This series of salaries may be grouped somewhat without any sacrifice of accuracy by merely tabulating the number of teachers who receive each amount. Such a tabulation is presented at the side. The column showing how many salaries of each amount there are is headed *f*, since each entry therein is called a *frequency*. *N* is the usual symbol employed for the total number of cases.

| *Salary* | *f* |
|---|---|
| 1900 | 1 |
| 1800 | 2 |
| 1750 | 1 |
| 1700 | 1 |
| 1650 | 1 |
| 1600 | 3 |
| 1550 | 2 |
| 1500 | 5 |
| 1450 | 5 |
| 1400 | 4 |
| 1350 | 6 |
| 1300 | 3 |
| 1250 | 3 |
| 1200 | 2 |
| 1150 | 2 |
| 1100 | 2 |
| 1050 | 2 |
| 1000 | 3 |
| $N =$ | 48 |

Although the tabulation just given is distinctly easier to comprehend and employ in computation than the list of 48 scores, even when they are in order, yet it may be desirable to condense the scores still more. For most purposes they can be grouped into classes 100 in width without too great loss of accuracy. This has been done and the resulting tabulation given on page 20. In this the entries in the first column are not, as in the previous one, the exact salaries, but are the *lower limits*, abbreviated *l* or sometimes *ll*, of the classes within which they fall. The dash after each indicates that the class extends all the distance up to, but not including, the next lower limit. For

example, the 975–class contains all cases from 975 up to but not including 1,075. In other words, its *upper limit*, abbreviated *u* or *ul*, is 1,074.99.

| $l$ | $f$ |
|---|---|
| 1875– | 1 |
| 1775– | 2 |
| 1675– | 2 |
| 1575– | 4 |
| 1475– | 7 |
| 1375– | 9 |
| 1275– | 9 |
| 1175– | 5 |
| 1075– | 4 |
| 975– | 5 |
| $N =$ | 48 |

The usual method of making such a tabulation from individual items is to determine what classes are to be used, list their lower limits, and record by tally marks, with a horizontal or diagonal line through each four to make a block of five, how many fall in each class. Figures are then substituted for the marks.[3] Very few workers are so accurate but that, in order to insure correctness, they should tabulate a series of scores twice and secure agreement between the two tabulations before accepting the result.

There are several principles to follow in deciding what classes to use in a tabulation. These principles are intended to provide for as great economy as possible in manipulation of data without serious sacrifice of exactness. Sometimes they conflict, thus rendering compromise or neglect of one principle necessary.

As a preliminary to discussing these principles, some terms will be defined. A *class* or *group* is one of the divisions into which the obtained range of values is divided according to size or, in the case of attributes, to some other characteristic. The *lower limit* of a class is the lowest possible value of any case that falls within it, and the *upper limit* is the highest. Since the upper limit of any class differs from the lower limit of the next class by an infinitesimal amount, that lower limit is generally substituted for the upper limit of the next lower class when the latter is employed in computation. Sometimes the term *boundary* is used instead of *limit*.

The distance covered by a class, from its lower limit to the lower limit of the next class, is best called the *class interval* or simply the *interval*. Sometimes *class width* or just *width* is

---

[3] In practice, it is frequently convenient to make the tally marks with a soft pencil, then write in the figures in ink, and finally erase the tally marks to leave a clean, neat tabulation.

employed. Frequently the word *interval* is also used as synonomous with *class*. In addition, *step* is employed in both meanings. Usually the context makes clear which meaning should be attached to *interval* or *step*. Thus such expressions as "the 1200-interval," "the 25-step," and so on evidently mean the same as *class*, whereas "the interval is 100" and "the step is 10" refer to the width of classes, that is, to the distance between the limits.

The usual recommendation as to the number of classes in a distribution is that it be between ten and twenty. The use of less than ten is liable to result in too great loss of accuracy, whereas that of more than twenty does not permit ready comprehension of the whole distribution or sufficiently reduce the work of computation. Ten and twenty should not be regarded as absolute, however; sometimes less than ten or more than twenty classes may well be employed. Usually a large number of cases should be tabulated in more classes than a small number.

Ease of manipulation increases if intervals are round, familiar numbers, such as 5, 10, 20, or 100, rather than such as 7, 11, 22, or 130. Moreover, tabulation is easier if intervals are so chosen that each lower limit differs from the adjacent ones in as many figures as possible. For example, an interval of 20, with accompanying lower limits of 0, 20, 40, 60, and so on, permits easier tabulation than one of 25, with lower limits of 0, 25, 50, 75, and so on. In the former case, the digits in units' place need not be watched at all; in the latter, they must be if the item falls in the 20's, the 70's, the 120's, and so on.

The suggestion just given should not be followed when doing so would not produce frequency distributions that represent the original data with sufficient accuracy. Since the mid-point of a class is taken as the average value of the cases included in the class, intervals should be chosen with this in mind. The chief point to watch is that if there is a strong tendency for the original data to differ from one another by a uniform amount, the interval should be an integral multiple of this difference and the lower and upper limits should be so chosen that each falls halfway between clusters of cases. For example, in the

salary tabulation used before, differences are 50, or multiples thereof; hence the interval employed should be 50, 100, or some other integral multiple of 50 and the limits should end in 25 or 75. Since, as the data were tabulated, 975, 1,075, and so on are the lower limits, the mid-points are 1,025, 1,125, and so forth, and these are the averages of 1,000 and 1,050, 1,100 and 1,150, and other values ending in 00 and 50, as they should be.

If the smallest salary increment were 40 rather than 50, and larger differences were integral multiples of 40, then a class interval such as 80 should be employed, with limits 20 below the next highest or above the next lowest salary. For example, if the ten smallest salaries were 1,200, 1,200, 1,160, 1,160, 1,120, 1,080, 1,080, 1,040, 1,000, and 1,000, and the total range such that an interval of 80 seems satisfactory, the lowest class should begin at 980, the next at 1,060, and so on. Thus the mid-points would be 1,020, 1,100, and so forth—in each instance the average of the two sizes of salary included in the class.

When the principle just stated is not violated, lower and upper limits may well be selected so as to be integrally divisible by the interval. Thus, if the interval is 10, limits such as 10, 20, 30, 40, and so on are generally preferable to 15, 25, 35, and so on; if the interval is 50, limits such as 200, 250, and 300 are usually better than those of 225, 275, and 325; and similarly in other cases.

The class limits of a distribution should be so written as to show the accuracy of the measures tabulated. This is generally best accomplished by carrying them to as many decimal places as were used for the original measures. For example, such lower limits as 10., 20., 30., 40., and so on indicate that the original data were accurate to the nearest unit, whereas 10.00, 20.00, 30.00, and so on indicate accuracy to the nearest hundredth.

Sometimes distributions of data are such as to provide exceptional tabulation difficulties. For example, if in addition to the salaries of the forty-eight teachers there is included also that of the superintendent, 3,750; that of the high-school

principal, 2,700; and those of two elementary principals at 2,100 and 2,000, to tabulate all in intervals of 100 would require twenty-eight intervals with lower limits ranging from 975 up to 3,675 and would make a lengthy distribution. On the other hand, to employ intervals of 200 or more would produce satisfactory length, but would condense the bulk of the data too much. The best solution is probably that shown at the side—to employ intervals of 100 but not to insert them where several adjacent ones include no cases, that is, have frequencies of 0. Instead two or three dots, in vertical arrangement, are inserted to indicate where each omission occurs, thus resulting in a tabulation such as that shown.

| *Salary* | *f* |
|---|---|
| 3675– | 1 |
| ⋮ | |
| 2675– | 1 |
| ⋮ | |
| 2075– | 1 |
| 1975– | 1 |
| 1875– | 1 |
| 1775– | 2 |
| 1675– | 2 |
| 1575– | 4 |
| 1475– | 7 |
| 1375– | 9 |
| 1275– | 9 |
| 1175– | 5 |
| 1075– | 4 |
| 975– | 5 |
| $N$ = | 52 |

Occasionally an even more exceptional set of data almost or absolutely requires the use of intervals of unequal width. Otherwise there is no possible way to avoid either too great condensation or very cumbersome length. General income statistics are an excellent example of such a situation. If any reasonably small number of equal intervals is used, the lowest one extends so high as to include many more than 99 per cent of the total number of cases. Therefore several widths of intervals are employed. A possible set of lower limits is given at the left. In this case the higher class has an indefinite upper limit. Another common example in many states or other sections is high-school enrollments, the large majority of which are quite small but some of which extend into the thousands.

| *Income* |
|---|
| 1,000,000– |
| 100,000– |
| 75,000– |
| 50,000– |
| 25,000– |
| 20,000– |
| 15,000– |
| 10,000– |
| 5,000– |
| 4,000– |
| 3,000– |
| 2,000– |
| 1,000– |
| 0– |

In the frequency distributions so far presented only the lower limits are given. Some persons prefer to insert the upper limits as well, presumably for the sake of clarity, but to the present writer there does not appear to be sufficient gain in so doing to warrant the practice. If it is done, the limits of the highest three classes in the salary tabulation of the forty-eight teachers appear as at

| |
|---|
| 1875.–1974.99 |
| 1775.–1874.99 |
| 1675.–1774.99 |

| |
|---|
| 1875–1975 |
| 1775–1875 |
| 1675–1775 |

the left. A somewhat similar practice that should never be followed is to write them as shown at the right. This form does not make clear where a salary of 1,775, for example, should be tabulated.

Sometimes instead of using limits to define the classes, workers give their mid-points as illustrated here. Such mid-points are called *class marks*, *class types*, *class indices*, or *face values*. The practice is statistically correct if properly understood, but both renders tabulation more difficult and is liable to misunderstanding.

1925
1825
1725

Before closing the discussion of tabulation and classification let us carry through an actual example. For this purpose the following set of total scores of a group of pupils upon a series of short tests may be employed: 214, 190, 194, 192, 267, 244, 219, 228, 195, 274, 248, 368, 219, 244, 239, 192, 301, 261, 234, 258, 215, 247, 236, 396, 307, 188, 242, 266, 232, 224, 257, 243, 326, 221, 311, 200, 233, 300, 244, 342, 237, 372, 184, 331. Inspection of these scores reveals that the smallest is 184; the largest, 396; hence the range is 396 − 184 = 212. Since they appear to be fairly well distributed over the whole range and no other complicating or exceptional condition is apparent, an interval of 20, which will make eleven classes, with limits into which it is integrally divisible, seems satisfactory. Such limits are therefore set up and the scores tabulated by tally marks, as indicated. For the first score, 214, a tally is placed after 200–; for the next, 190, a tally is placed after 180–; and similarly for all forty-four. This should then be repeated and the two sets of tallies compared to see if they agree. If so, they are probably correct and may be accepted. Figures should then be substituted for them, either at the side, as shown here, or wherever else may be convenient, and the tabulation is ready for use.

| *Score* | *f* |
|---|---|
| 380–1 | 1 |
| 360–11 | 2 |
| 340–1 | 1 |
| 320–11 | 2 |
| 300–1111 | 4 |
| 280– | 0 |
| 260–1111 | 4 |
| 240–~~1111~~ 1111 | 9 |
| 220–~~1111~~ 1111 | 9 |
| 200–~~1111~~ | 5 |
| 180–~~1111~~ 11 | 7 |
| | $N = 44$ |

## Assumed location of grouped cases

The assumed location of grouped cases is quite important in

several connections. Since cases lose their original exact values when grouped in a distribution, some assumption as to their values must be made as a basis of many of the computations and interpretations involving them. Although another assumption would often be more accurate, the one regularly made is that the cases in a class are symmetrically and equally distributed throughout that class. This means that each class interval is divided into as many equal parts or distances as there are cases in that class and one case assumed to be at the mid-point of each part.

The assumption just stated may be illustrated by Figure 1,

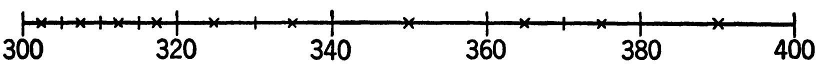



**Fig. 1. Graphic Representation of Assumed Distribution of Cases in a Grouped Distribution**

which represents graphically the five classes at the top of the distribution used previously, with the assumed position of each case shown. Since the class extending from 300 up to 320 has a frequency of 4, it is divided into 4 equal parts, and at the mid-point of each is an x, to represent the case assumed to be there. Since one-fourth of 20, the interval, is 5, the first part extends from 300 to 305 and the case at its mid-point is at 302.5. The second part extends from 305 to 310, so the second case is at 307.5, and similarly the others are 312.5 and 317.5. The 320-class contains 2 cases; hence it is divided into 2 parts covering from 320 to 330 and 330 to 340, respectively, and the cases are at their mid-points, 325 and 335. The 340-class has but 1 case, which therefore is at its mid-point, 350. The 2 cases in the 360-class are at 365 and 375, and the 1 in the 380-class is at 390.

In the instances just presented it is easy to determine the assumed value of each case, but when the class frequency is large it is difficult to do so by mere inspection. The values found in the last paragraph show that the first case in a class is always $\frac{1}{2f}$ of the interval above the lower limit and that each

successive case is $\frac{1}{f}$ of the interval larger than the case just below it. In formula form, the assumed value of any case in a class is $l + \frac{n - .5}{f} i$, which may also be written $l + \frac{2n - 1}{2f} i$. In this $l$ is the lower limit of the class involved, $f$ is its frequency, $i$ is the interval, and $n$ is the ordinal number, or position within the class of the case the value of which is desired. For the value of the second case in the 300-class this formula gives $300 + \frac{2. - .5}{4} 20 = 307.5$, the result already secured by inspection. As another example, the assumed value of the third case in the 220-class of the distribution of scores is $220 + \frac{3. - .5}{9} 20 = 225.56$.

For the first case in a class it is convenient to use a simplified form of the formula. Substituting 1 for $n$ gives

$$l + \frac{2 \times 1 - 1}{2f} i = l + \frac{i}{2f}$$

as the value of the first case. For example, the first case in the 200-class is $200 + \frac{20}{10} = 202$. Similarly the formula for the last case is $u - \frac{i}{2f}$, $u$ being the symbol for upper limit. Thus the last case in the 240-class, for example, is $260 - \frac{20}{18} = 258.89$. It is also convenient to remember that if the frequency in a class is an odd number the middle case always comes at the mid-point of the class. For example, the fifth or middle one of the 9 cases in the 220-class is 230.

## Meaning of numerical data

The numerical expressions of scores and other data do not always have the same meaning. There are two chief interpretations. In some—probably most—cases, a score or other datum of a given value means at least that much, but not so much as the next higher value. For example, a score of 46

on a spelling test regularly means 46 words spelled entirely correctly, with perhaps some others partially right, but not 47 correct. The same is true for scores on almost all varieties of short-answer tests, both standard and informal.

The other common but probably less frequent type is that in which a score means nearer the given value than to the next one below or above it. For example, the directions for scoring specimens of pupils' handwriting by the Ayres Handwriting Scale, which has samples at intervals of 10 from 20 to 90, provide that each be given the value of the sample which it most resembles. In other words, specimens rated between 25 and 35 are given 30, those between 35 and 45 are given 40, and so on. The same interpretation is usually given to per-cent marks on essay examinations, to intelligence quotients, and to various other types of data.

In the cases of some varieties of data there is no general agreement in practice as to which is the meaning of the data. For example, an age of 11 may mean 11 but not 12 years, or it may mean nearer to 11 than to either 10 or 12.

It is important that whoever is concerned with a set of data know which interpretations and results are appropriate for them. Misinterpretations and consequent errors frequently result because workers fail to make clear which type of data they are employing. Often the source of confusion has been failure to give class limits n such a way as to indicate the actual meaning of the data. If they are of the second type, such that a given score means nearer to the stated value than to those just below and above it, the class limits should be stated so as to make this fact evident. Thus for scores on the Ayres Handwriting Scale lower l mits of 15, 25, 35, and so on, rather than of 20, 30, 40 and so forth, should be employed. Similarly for ordinary per-cent marks, since one of 60 means from 59.5 to 60.5, one of 61 from 60.5 to 61.5, and so forth, lower limits for intervals of 5 should be 59.5, 64.5, 69.5, and so on, rather than 60, 65, 70, and so on.

The safest practice to follow in interpreting data is to take the given limits as the true limits unless it is definitely known that they are not. In other words, mid-points should regularly

be taken halfway from one lower limit, as given, to the next, unless it is certain that they should be taken otherwise.

## Cumulative frequency distributions

The type of frequency tabulation or distribution described and employed above is the most usual variety employed. It is, however, sometimes helpful to make use of *cumulative distributions*. The entries or frequencies in such distributions show the number of cases in and below, or in and above, each class instead of just the number in it. Such a tabulation is commonly made from an ordinary one of the kind presented above. The frequencies therein are summed continuously from one end or the other, but more often from the lower end, to give the cumulative frequencies. Such a tabulation is illustrated in Table I. The first two columns merely repeat the distribution found on p. 24. The third contains the cumulative frequencies obtained by successively adding the ordinary frequencies from the bottom up. The cumulative frequency in the lowest class is always the same as its ordinary frequency, in this instance 7. That in the next or 200-class is 12, the sum of its frequency, 5, and the cumulative frequency, 7, of the class below. The cumulative frequency in the 220-class is 21,

TABLE I

CUMULATIVE FREQUENCY DISTRIBUTION, CUMULATED FROM THE BOTTOM UP

| Score | *f* | *cum f* | *cum %* |
|---|---|---|---|
| 380– | 1 | 44 | 100.0 |
| 360– | 2 | 43 | 97.7 |
| 340– | 1 | 41 | 93.2 |
| 320– | 2 | 40 | 90.9 |
| 300– | 4 | 38 | 86.4 |
| 280– | 0 | 34 | 77.3 |
| 260– | 4 | 34 | 77.3 |
| 240– | 9 | 30 | 68.2 |
| 220– | 9 | 21 | 47.7 |
| 200– | 5 | 12 | 27.3 |
| 180– | 7 | 7 | 15.9 |
| $N =$ | 44 | | |

which is its frequency, 9, plus 12, the cumulative frequency of the 200-class. Similarly the others are found, finally that of the highest class being equal to $N$, here 44.

The last column is not necessary, but is often useful. It contains the cumulative frequencies expressed as per cents of the total number of cases. In other words, each entry is the corresponding cumulative frequency divided by 44. It may also be obtained by finding $\frac{100}{N}$ and multiplying by the cumulative frequency. Here $\frac{100}{N} = \frac{100}{44} = 2.273$. Seven times this $= 15.9$, the first cumulative per cent; $12 \times 2.273 = 27.3$, the next, and so on.

In interpreting such a cumulative table readers should note that since the cumulative frequency for each class represents the number of cases in and below that class, it signifies that there are that many cases below the lower limit of the next higher class. Thus the cumulative frequency of 7 for the 180-class indicates that 7 cases are below 200, that of 12 for the 200-class that 12 are below 220, that of 21 for the 220-class that 21 are below 240, and so on. The same interpretation, of course, applies to the cumulative per cents as well as to the frequencies.

TABLE II

Cumulative Frequency Distribution, Cumulated from the Top Down

| Score | $f$ | *cum f* |
|---|---|---|
| 380– | 1 | 1 |
| 360– | 2 | 3 |
| 340– | 1 | 4 |
| 320– | 2 | 6 |
| 300– | 4 | 10 |
| 280– | 0 | 10 |
| 260– | 4 | 14 |
| 240– | 9 | 23 |
| 220– | 9 | 32 |
| 200– | 5 | 37 |
| 180– | 7 | 44 |
| | $N = 44$ | |

Although frequencies are cumulated from the top down much less often than from the bottom up, the procedure is illustrated in Table II. The cumulative frequency of the highest class is 1, the same as its ordinary frequency. That of the next highest is 3, from $1 + 2$; that of the next is 4, from $3 + 1$; and so on. Finally that of the lowest class is equal to $N$, 44. In this case each cumulative frequency indicates the number of cases above the lower limit of its class. Thus there is 1 case above 380; there are 3 above 360, 4 above 340, 6 above 320, and so on. Cumulative per cents may also be found here, but they have not been computed and included in the table.

## EXERCISES AND PROBLEMS

**1.** Which of the following are usually best dealt with as attributes and which as variables?

(*a*) Places of birth
(*b*) Grade enrollments
(*c*) Amounts of tardiness
(*d*) Subjects failed by pupils
(*e*) Colleges attended
(*f*) Ratings of textbooks
(*g*) Professional training of teachers
(*h*) Nationalities of pupils
(*i*) Quality of handwriting
(*j*) Causes of absence

**2.** Tabulate the following ages in a frequency distribution:
15–6, 16–1, 14–11, 17–2, 14–8, 15–10, 16–4, 13–9, 14–4, 17–5, 14–10, 16–3, 17–9, 18–4, 16–9, 14–10, 16–6, 17–0, 16–7, 15–5, 17–1, 18–7, 14–6, 19–0, 14–2, 15–8, 16–7, 17–4, 17–6, 17–8, 14–9, 15–1, 17–8, 16–10, 13–5, 15–10, 14–7, 16–4, 18–2, 15–11, 19–3, 14–8, 16–0, 17–4, 15–3, 17–0, 15–9, 16–2, 15–1, 15–11, 16–8, 18–4.

**3.** Tabulate the following school enrollments in a frequency distribution:
71, 18, 243, 95, 51, 26, 76, 141, 428, 48, 36, 78, 122, 341, 50, 16, 65, 108, 254, 29, 94, 35, 12, 33, 66, 209, 37, 82, 8, 136, 81, 841, 59, 30, 74, 76, 34, 40, 70, 181, 21, 389, 172, 13, 88, 37. 62, 94, 100, 11, 42, 31, 64, 383, 55, 41, 17, 29, 36, 23, 75, 162, 198, 68, 44, 21, 71, 40, 86, 184, 645.

**4.** Tabulate the following ratings of school buildings in a frequency tabulation:
863, 794, 623, 852, 921, 781, 724, 605, 766, 841, 702, 675, 773, 841, 809, 730, 934, 622, 556, 687, 752, 834, 764, 825, 677, 842, 738, 872, 696, 571, 832, 768, 718, 864, 758, 660, 683, 819, 736, 628, 904.

**5.** Tabulate the following intelligence quotients in a frequency tabulation:
88, 102, 95, 79, 106, 89, 97, 99, 124, 110, 85, 98, 106, 92, 113, 83, 94, 99, 108, 76, 93, 145, 104, 95, 101, 80, 100, 73, 97, 109, 125, 94, 116, 87, 93,

107, 104, 116, 93, 81, 114, 95, 87, 104, 126, 94, 109, 103, 99, 107, 109, 87, 94, 90, 106, 119, 127, 107, 103, 74, 88, 93, 77, 104, 94, 109, 98, 103, 117, 86, 93, 97, 91, 106, 115, 158, 93, 88, 104.

**6.** Make the cumulative frequency distributions, both up and down, of the data in Exercises 2, 3, 4, and 5. Also find the cumulative per cents for each.

**7.** Find the assumed value of each indicated case in the accompanying distribution.

(*a*) The first case in the 50-class.
(*b*) The second case in the 65-class.
(*c*) The seventh case in the 60-class.
(*d*) The one case in the 45-class.
(*e*) The highest case in the 70-class.

| *Score* | *f* |
|---|---|
| 75– | 2 |
| 70– | 4 |
| 65– | 5 |
| 60– | 9 |
| 55– | 8 |
| 50– | 2 |
| 45– | 1 |
| 40– | 1 |
| | $N = 32$ |

**8.** Find the assumed value of each indicated case in the accompanying distribution.

(*a*) The third case in the 500-class.
(*b*) The lowest case in the 750-class.
(*c*) The eighth case in the 550-class.
(*d*) The fifth case in the 650-class.
(*e*) The fourth case in the 400-class.

| *Score* | *f* |
|---|---|
| 850– | 1 |
| 800– | 0 |
| 750– | 6 |
| 700– | 12 |
| 650– | 18 |
| 600– | 20 |
| 550– | 15 |
| 500– | 7 |
| 450– | 2 |
| 400– | 5 |
| | $N = 86$ |

## REFERENCES

Guilford, J. P., *Fundamental Statistics in Psychology and Education.* Pages 64–67. New York: McGraw-Hill Book Company, Inc., 1942.

Kenney, John F., *Mathematics of Statistics.* Part I, Chap. I. New York: D. Van Nostrand Company, Inc., 1939.

Mode, Elmer B., *The Elements of Statistics.* Chap. IV. New York: Prentice-Hall, Inc., 1941.

Monroe, Walter S. and Engelhart, Max D., *The Scientific Study of Educational Problems.* Pages 66–67. New York: The Macmillan Company, 1926.

Walker, Helen M., *Elementary Statistical Methods.* Pages 23–36, 43–46. New York: Henry Holt & Company, Inc., 1943.

## CHAPTER III

# Graphic Representation of Frequency Distributions

### The coordinate axis system and frequency curves

Although most of the graphic representations of frequency distributions belong in only one quadrant—the first—of the *coördinate axis system,* the worker in this field should have at least an elementary understanding of this system, so widely used in mathematical work. Figure 2 illustrates it. It consists of two perpendicular lines, the horizontal one of which

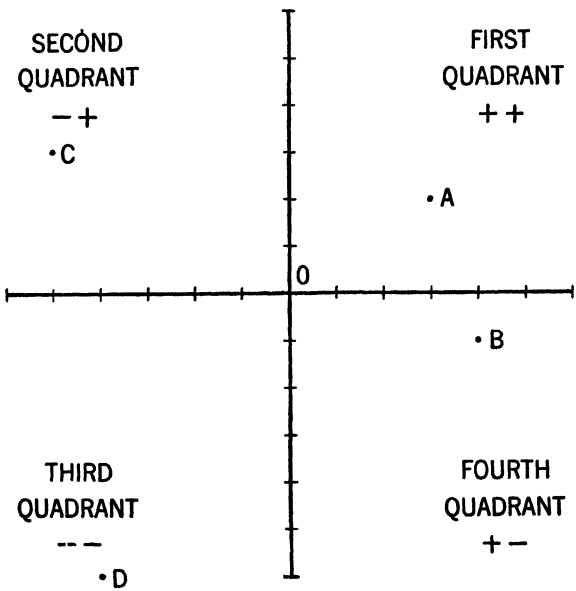



**Fig. 2. The Coördinate Axis System**

is known as the *X-axis* and the vertical one as the *Y-axis*. Their point of intersection is the *origin*, abbreviated *O*, and is the zero point from which horizontal and vertical distances are measured. Horizontal or $X$ distances to the right of the origin are positive, those to the left negative. Vertical or $Y$ distances above the origin are positive, those below it negative.

The four divisions of space formed by the two axes are known as *quadrants*. They are numbered ordinally as shown on Figure 2. The + and − marks in them indicate the signs of points located in them, the first mark being the sign of their $X$ distance and the second of their $Y$ distance. Thus in the first quadrant both $X$ and $Y$ distances or values are positive, in the second $X$ is negative and $Y$ positive, and so on.

The most frequent use of the axis system is to describe points by their location. This is done by giving two figures. The first of these, called the *abscissa*, shows the $X$ distance, and the second, called the *ordinate*, shows the $Y$ distance. For example, the location of point A is commonly given thus: 3,2, which means that it is 3 units to the right of the $Y$-axis and 2 units above the $X$-axis. Similarly point B $(4, -1)$ is 4 units to the right and 1 below; C $(-5, 3)$ is 5 units to the left and 3 above, and D $(-4, -6)$ is 4 units to the left and 6 below.

When, as is very common, the two variables to be represented graphically are the amount of a trait or characteristic and the frequency in each class thereof, regular practice is to use the $X$ or horizontal direction to represent the amount of the characteristic and the $Y$ or vertical direction for the frequencies. Since there are rarely cases with less than none of the characteristic concerned and never frequencies below 0, both variables regularly have positive values and the graph representing them lies in the first quadrant.

To represent a frequency distribution graphically, one should construct horizontal and vertical scales divided into units of such sizes as make its significant features readily apparent. In general the width and height of the graph may well be about equal. The scale for the amount of the characteristic should begin about one interval below the lowest case and extend about one above the highest. If its lower end is

not at 0, there should be a short broken line extended to the 0 point. The vertical or frequency scale, ordinarily erected at the 0 point, should be as high as needed, or slightly higher, to represent the largest frequency. Both scales should be clearly labelled. These and other desirable features of graphs will be illustrated in the next few figures. The three common forms of the ordinary frequency curve or graph—the histogram or column diagram, the polygon, and the smooth frequency curve—will be presented in the following sections.

## The histogram or column diagram

The *histogram* or *column diagram* is composed of a series of rectangles, each one interval in width and of proper height to correspond to the frequency of the interval it represents. The lines dividing one rectangle from another are generally omitted, but may be drawn in to emphasize the division into classes. Figure 3 contains such a graph. Its height above each class is

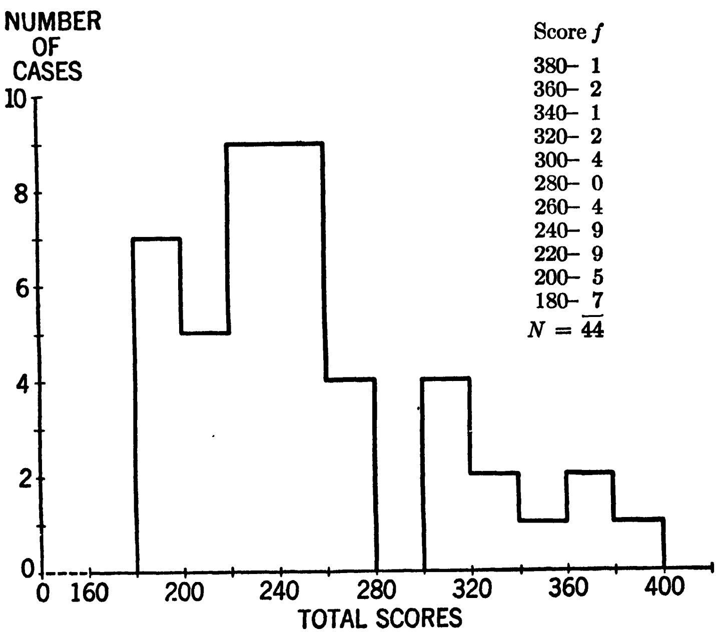



**Fig. 3. Histogram or Column Diagram Representing the Data Shown at the Right**

that indicated by the corresponding number of cases, 7 above the 180-class, 5 above the 200-class, and so on.

The easiest way to construct a histogram, after choosing and laying off the two scales in convenient dimensions, is to draw all the lines running in one direction and then all those in the other. Which set is drawn first is immaterial.

In rare instances histograms are presented with axes reversed, so that the scale for whatever is measured appears vertically and that for the frequencies horizontally. An adaptation of this that is convenient in the preparation of typed copy is the use of as many *x*'s, *o*'s, or other symbols as there are cases in each class. The result of this for the same distribution represented by the histogram is shown at the right.

```
380-x
360-xx
340-x
320-xx
300-xxxx
280-
260-xxxx
240-xxxxxxxxx
220-xxxxxxxxx
200-xxxxx
180-xxxxxxx
```

An advantage of the histogram over both the frequency polygon and the smooth frequency curve is that it can readily be employed to show individual cases. For this purpose the rectangle above each class interval is divided into as many sections, one above the other and usually square, as there are cases therein. Either one of two facts pertaining to each case, or both, may be shown. The unit sections may be labelled by number, letter, or otherwise so that the one representing each case can be identified,[1] or the exact value of each case may be inserted, or both.

Figure 4 illustrates such a use of a histogram. It represents the scores of 32 pupils upon a 200-element test. The exact scores are shown, with those in each class arranged in order from lowest up to highest. Thus one pupil made 102, one 105, and so on up to one who scored 185. If each pupil has been informed as to his score, there is no need for using identifying labels.

In such a graph no scale of frequencies is necessary, since the cases can readily be counted. If, as here, the actual scores are given, the base-line scale does not need to be labelled. If they are not, it should be, just as for an ordinary histogram.

[1] When pupils' scores are presented thus, it is usually best to label them so that only the pupil concerned and the teacher knows whose score each is.

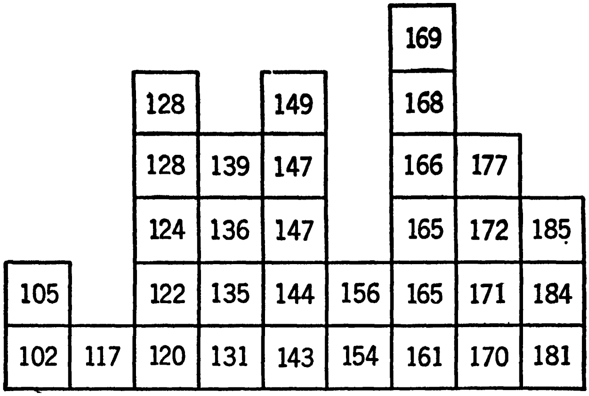



**Fig. 4. Histogram Divided Into Unit Sections, Each Representing an Individual Case**

## The frequency polygon

The *frequency polygon* is an angular curve formed by connecting with straight lines a series of points, one above the mid-point of each interval at a height corresponding to the frequency therein. At each extreme a line is drawn from the point above the last class to the base line at the mid-point of the next class below at the bottom, and above at the top, of the distribution.

Figure 5 illustrates this type of graph. The line bounding it begins on the base line at 170, the mid-point of the class just below the lowest class in the distribution. From that point it is drawn to another at a height of 7 above 190, the mid-point of the 180-class; thence to one at 5 above 210, the mid-point of the 200-class; thence up to a height of 9; and so on until it ends on the base line at 410, the mid-point of the next class above the last one. Since there are no cases in the 280-class, it comes down to and rises from the base line at 290, the mid-point of that class.

A frequency polygon does not include above the base-line distance representing each class an area that exactly corresponds to the frequency of that class. For this reason it does not show the frequency in each class so readily as the histogram does. Its total area, however, is equal to that of the histogram, and it so distributes that area as probably to be more nearly in

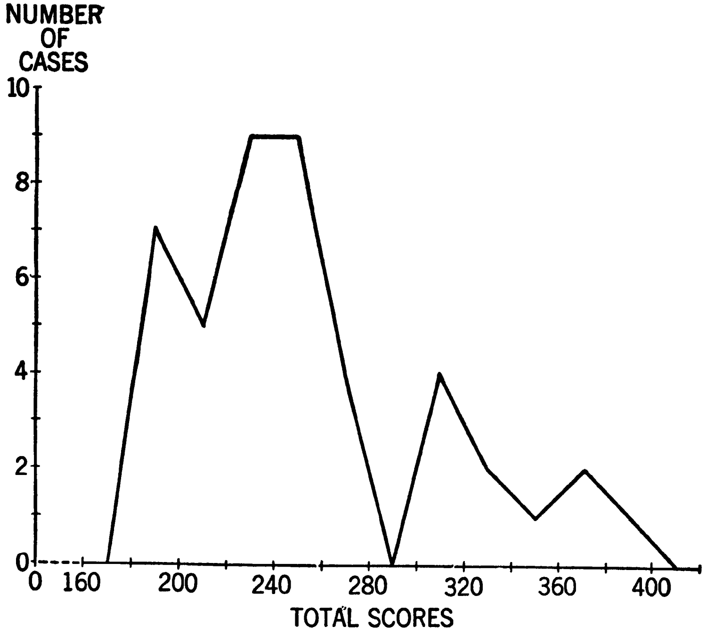



**Fig. 5. Frequency Polygon Representing the Same Data as the Histogram in Figure 3**

accord with the true distribution of the individual data than the histogram is. This is even more likely to be true if the actual data employed are a sample of a large number which the graph is to represent.

## The smooth frequency curve

The *smooth frequency curve* is the most graceful and artistic form of the three varieties, but it is also the most difficult to draw well. It is less accurate than the frequency polygon in portraying the number of cases in each class, but more likely to approach the true shape of the distribution of single data or of a larger number from which a sample was selected. Such a curve for the same data as used for the others is shown in Figure 6.

To draw such a curve, points are located just as for a frequency polygon and a smoothly flowing curved line is drawn through them instead of a series of straight lines.

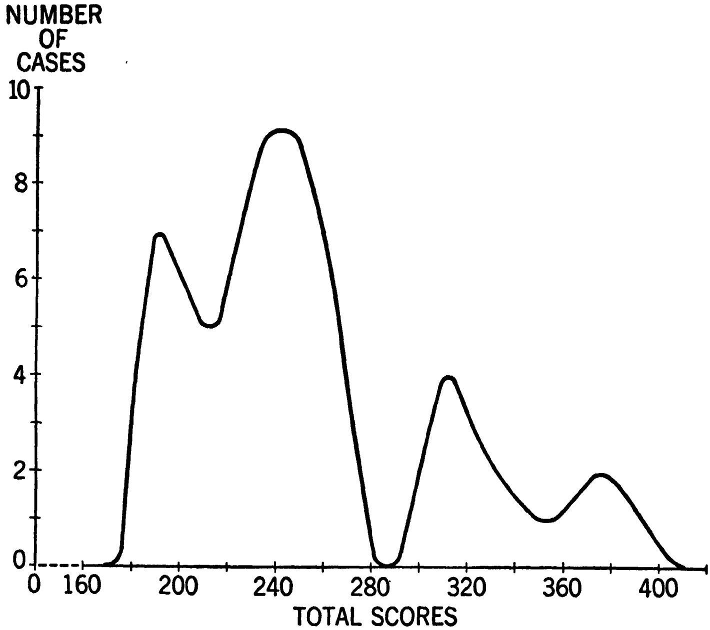



**Fig. 6. Smooth Frequency Curve Representing the Same Data as the Graphs in Figures 3 and 5**

## Smoothing

The *method of smoothing,* also called that of *graduating* or of *moving* or *rolling averages,* is sometimes applied to a distribution, or the curve that represents it. Its purpose is to eliminate minor fluctuations due to small frequencies, irregular movements in time series, poor sampling, short time cycles, or other causes of unrepresentative sampling or variable errors. For example, if 50 cases are taken as a sample from 1,000, the fraction of the 50 falling in each class will probably vary somewhat erratically from the corresponding fraction of the whole 1,000. If the sample is properly smoothed, the result will probably more nearly represent the whole 1,000 than if it is not. Along with its desirable effect, however, *smoothing* may reduce or even eliminate features of the distribution that are actual characteristics of the larger group of data and should be retained. Therefore it should be employed with caution. If one is in doubt, he probably should not smooth data.

The smoothing of a set of data with short time cycles or seasonal movements so as to reduce or eliminate them and thus make the long-time trend more evident is often desirable. For example, if we wish to study the increase in spelling ability from year to year and have monthly scores as our data, we can eliminate much of the variation from month to month and emphasize the general trend by proper smoothing.

The general principle of smoothing is to replace each frequency or value by a smoothed frequency obtained by averaging several that center at the class in question. An odd number of frequencies, most often three, is generally taken to be averaged. Sometimes the middle one is given double weight in determining the smoothed frequency. So doing reduces the amount of change from the original frequencies. The chief situation in which even numbers of frequencies are averaged is in connection with time cycles of even numbers of units. Thus to eliminate monthly cycles, twelve frequencies may be averaged. Since, however, it is more convenient to employ an odd number, thirteen may be used.

Smoothed frequency curves may be drawn in any of the three forms already presented. If a smoothed curve is to be compared with the curve of the original data, it should be drawn similarly to the latter.

To illustrate smoothing by threes, the usual method, Figure 7 presents the same data already employed, the smoothed distribution by threes, and the frequency polygons representing both. The solid line is the original polygon, the same as in Figure 5, and the broken line the smoothed one.

The first smoothed frequency, 6.33, is obtained by doubling the first frequency, 7, adding the next one, 5, and dividing by 3. In other words, $\frac{2 \times 7 + 5}{3} = 6.33$. The next is the average of the frequencies for the three successive classes, of which it represents the middle one. Hence it is $\frac{7 + 5 + 9}{3} = 7.00$, and so on for the others. The last, as the first, requires doubling the last original frequency, hence is $\frac{2 + 2 \times 1}{3} = 1.33$.

It is evident that the smoothed distribution and curve are distinctly less uneven, that is, have smaller changes from class to class, than do the original ones. If a larger number of classes is averaged for each smoothed frequency the smoothing

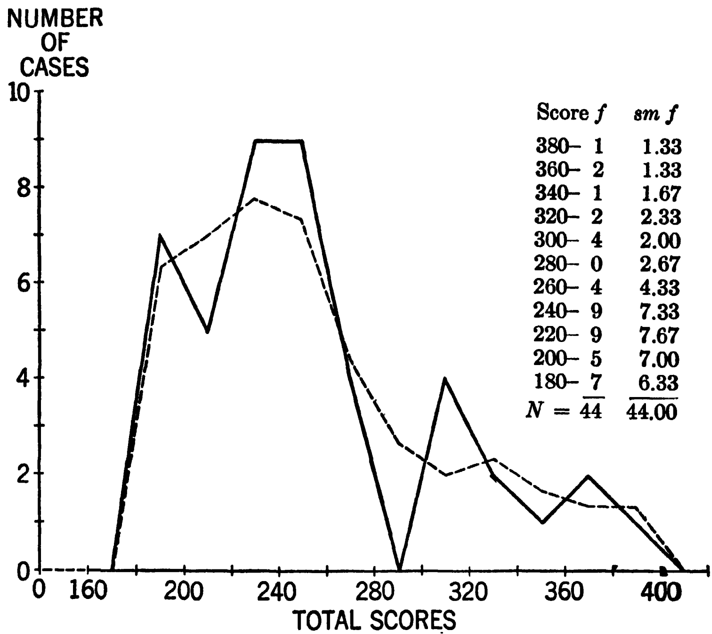



**Fig. 7. Original and Smoothed Frequency Polygons**

or levelling effect is still greater. Occasionally more than one smoothing is applied to the same data, which also results in greater reduction of differences from class to class.

It is possible to accomplish the same result by geometric or graphic means, that is, by drawing the original curve and then constructing the proper lines. The procedure is rarely employed, however, since for most persons it is more difficult than the method given.

## Cumulative frequency curves

Cumulative distributions may be represented by curves just as may ordinary distributions. Any of the three forms may be used, but in practice the polygon or angular form is almost

never employed, the rectangular or "stair-step" form is somewhat rare, and the smooth form decidedly the most frequent. Therefore the examples given will be in that form. Just as is the case for cumulative distributions, so cumulative curves or graphs may either be started from the bottom up, thus showing the number of cases below any specified point, or from the top down, showing the number above. The former is more common.

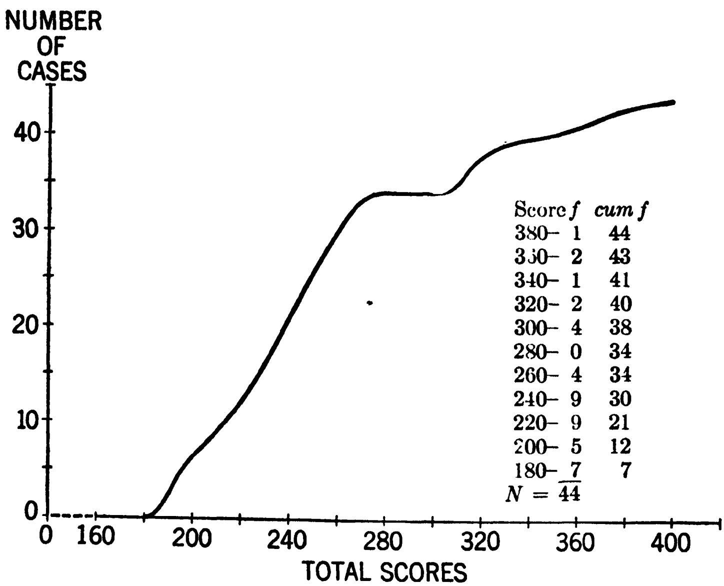



**Fig. 8. Cumulative Frequency Curve Showing Number of Cases Below Each Point**

Figure 8 presents a smooth curve cumulated from the bottom up for the same set of data already frequently employed in this volume. It leaves the base line at 180, the lower limit of the first class; rises to a height of 7 at 200, the lower limit of the second class; to 12 at 220; and so on up to 44 at 400. Thus it shows that no cases are below 180, 7 are below 200, 12 below 220, and so on.

Readers will note that the horizontal or score scale is the same in this figure as in the several others representing the

same data, but that the vertical or number of cases scale differs. Here it must provide for as many cases as are in the whole distribution, whereas in the others it needs to provide for only so many as the greatest frequency in any one class.

Since data are cumulated from the top down much less frequently than from the bottom up, no graph to illustrate such a cumulation is given. One for the same data just used would have scales the same as that in Figure 8. The curve would begin at a height of 44 above 180, be 37 high above 200, 32 high above 220, and finally reach the base line at 400. If it were plotted on Figure 8, the two curves would intersect directly above the median score, that is, the score or point on each side of which half of the scores lie.

## Ogives or percentile curves

There is some variation in practice in the use of the terms *cumulative curve*, *ogive*, and *percentile curve*. Some statisticians use them interchangeably, to refer to such a curve as that illustrated in Figure 8 and also to the same type of curve drawn with the two scales interchanged. Others employ *cumulative curve* as referring only to the type in Figure 8, and *ogive* and *percentile curve* for the variety with scales interchanged. The present writer favors this latter usage, hence will employ the two latter terms to refer to a graph representing cumulative frequencies with the trait measured shown on the vertical scale and the number of cases on the horizontal.

Ogives may be drawn for distributions cumulated in either direction, but their use for those cumulated from the top down is so rare as to be negligible. Moreover, the definition of a percentile is that it is a point with a certain per cent of the total number of cases below it, hence the term *percentile curve* or *percentile graph* can properly be applied only when the cumulation has been from the bottom up. Also, ogives may be drawn in any one of the three geometric forms of ordinary curves, but are usually smooth, occasionally rectangular, almost never angular.

Figure 9 presents the ogive for the same data already used. It shows just the same facts as does Figure 8, but with the

scales reversed. In addition to the scale showing the number of cases, a percentile scale is also laid out horizontally. It indicates where each tenth percentile comes, with vertical lines drawn up from each such point on the scale, as is frequently done. Thus there is a vertical line above the 10th percentile, wh'ch is the point below which 10 per cent of 44 or 4.4 cases lie;

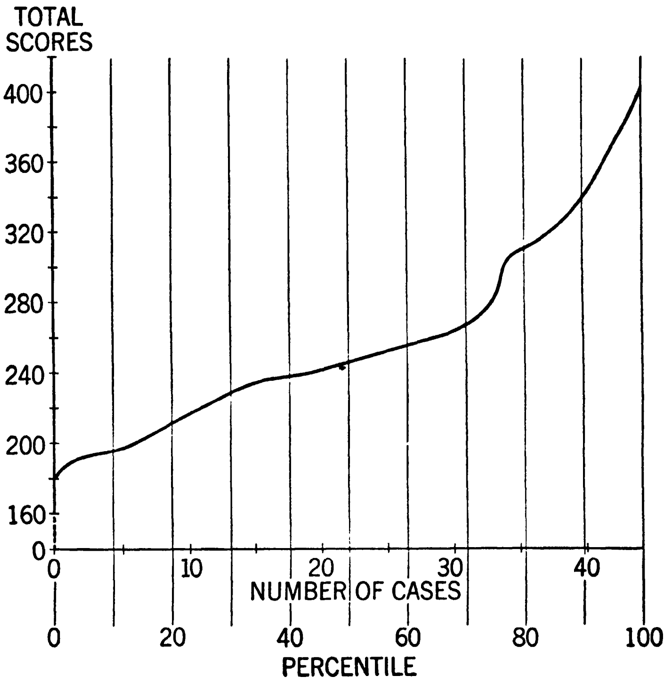



**Fig. 9. Ogive of Same Data Used in Figure 8 and Previous Figures**

another above the 20th percentile or the point below which 8.8 cases lie; and so on. Such vertical lines, by their intersections with the curve, assist in the determination of the score at which each percentile falls. Thus the 10th percentile line intersects the curve at a point even with approximately 192 or 193 on the vertical or score scale; the 20th percentile intersects it at a point corresponding to a score of about 207; the 30th at one corresponding to about 223; and so on. In other words, 10 per cent of the cases are below 192 or 193, 20 per cent

below about 207; and similarly for other percentiles. Sometimes horizontal lines are drawn to assist in determining at what score the percentiles fall, but this has not been done here.

In the example just given, each tenth percentile is shown, but there is no unanimity of practice in this respect. Sometimes fewer, and sometimes more, are shown. The 25th, 50th, and 75th are perhaps the most commonly indicated. In graphing test norms, a situation in which this type of curve is often employed, seven percentiles—the 5th, 10th, 25th, 50th, 75th, 90th, and 95th—are frequently shown.

In order to illustrate further the use of the ogive with test scores, Figure 10 has been prepared and inserted. It illustrates

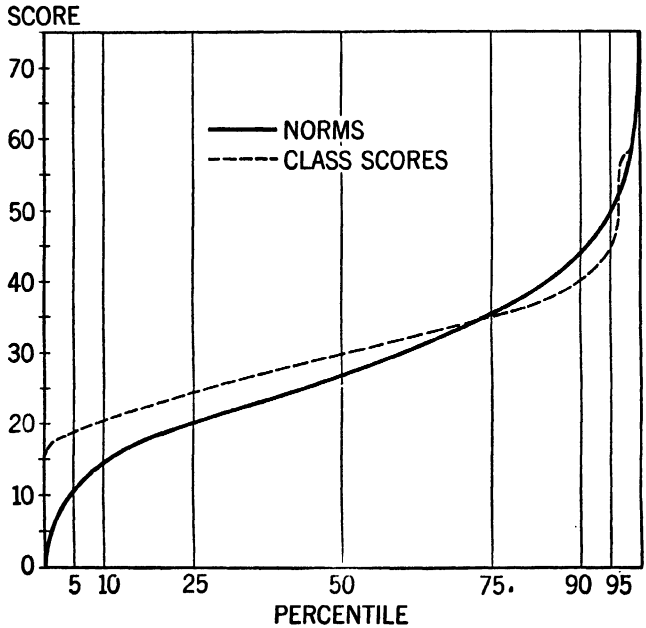



**Fig. 10. Ogive Illustrating Comparison of Scores of a Group with Norms for Data in Table III**

the comparison of scores made by a class of pupils with general norms. The solid line represents the norms, the broken one the scores of the class, as given in Table III. The actual numbers of cases are not shown on the figure. Since they differ,

TABLE III

SCORES SERVING AS BASIS OF NORMS AND THOSE FROM A SINGLE CLASS

| | Norms | | Class | |
|---|---|---|---|---|
| Score | $f$ | *cum f* | $f$ | *cum f* |
| 70– | 2 | 2372 | | |
| 65– | 4 | 2370 | | |
| 60– | 10 | 2366 | | |
| 55– | 19 | 2356 | 1 | 34 |
| 50– | 56 | 2337 | 0 | 33 |
| 45– | 78 | 2281 | 0 | 33 |
| 40– | 171 | 2203 | 2 | 33 |
| 35– | 241 | 2032 | 5 | 31 |
| 30– | 325 | 1791 | 7 | 26 |
| 25– | 438 | 1466 | 11 | 19 |
| 20– | 443 | 1028 | 6 | 8 |
| 15– | 328 | 585 | 2 | 2 |
| 10– | 177 | 257 | | |
| 5– | 71 | 80 | | |
| 0– | 9 | 9 | | |
| $N$ = 2372 | | | 34 | |

both should be there if either is and their presence would tend to complicate the graph unnecessarily. Instead, as is usually preferable when distributions containing different numbers of cases are represented on the same graph for comparisons such as this, per cents are used.

The positions of the two curves indicate not only the two distributions of scores, but also their comparative positions. Since at the 50th percentile the curve for the class is above that of the general norms, it is evident that on the whole the class scored high. Indeed, the whole lower 70 per cent of the class rank above the corresponding portion of pupils in general. The upper 30 per cent, however, are mostly below the norms. At both extremes the norms extend considerably farther than the class scores, thus indicating that the class is more homogeneous in achievement than are pupils in general.

## The normal frequency curve

Because approximations to the *normal frequency curve* result so often when biological—including educational—and many

other kinds of data are tabulated and graphed, the normal frequency curve is of high importance. Many terms are applied to it, among them being *normal*, *normal frequency*, *probability*, *error*, *biologic*, *Gaussian*, and *Laplacean*. It is symmetrical, bell-shaped, high at the center, decreasing in height fairly rapidly near the center, then more slowly, finally reaching the base line at an infinite distance from its center.

Figure 11 presents two normal frequency curves, drawn on the same horizontal scale but different vertical scales. Both

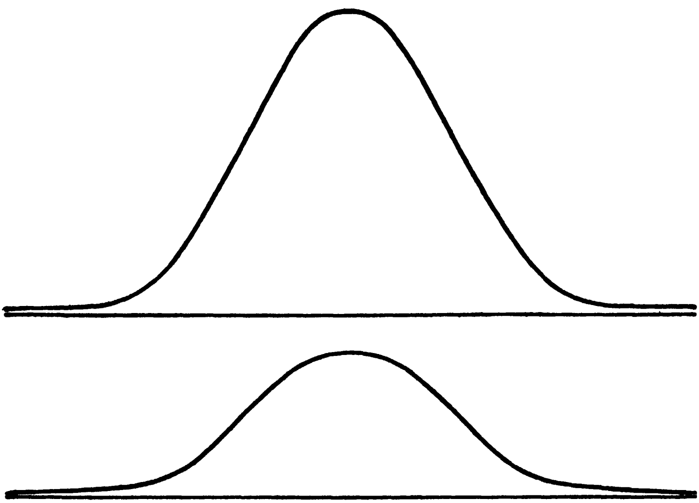


Fig. 11. Normal Frequency Curves

are shown to illustrate the fact that the scale may be chosen, according to the judgment of the person who constructs the curve, without robbing the curve of its normality. Each has been drawn to the points where it approaches the base line closely enough so that the space between can scarcely be shown.

The fact that approximations to normal distributions occur so frequently may lead workers to assume that almost every set of data should take such a shape and that if one does not something is wrong. In many instances this is true and failure to approximate normality is due to inadequate or nonrepresentative sampling, to presence of errors, or to use of a scale of

measurement with too crude units, but in others the true shapes of the distributions are not normal.

Not only do scores on most tests adapted in difficulty to the group tested, intelligence quotients of the general population, and many other varieties of data, approach normality, but the same is also true of variable errors, of matters of chance with equal odds, and of many other situations. The expansion of the binomial with the two terms equal yields an approximately normal distribution. The higher the power, the closer the approximation. Therefore such a matter of even chances as numbers of heads and tails resulting from tossing coins, for example, yields a distribution very nearly normal.

Although no use of the formula for the normal curve is made in this volume, it seems appropriate to give it. It is

$$y = y_0 e^{\frac{-x^2}{2\sigma^2}},$$

in which $y$ is any ordinate; that is, its height at any point; $y_0$ the maximum ordinate or height at the center; $e = 2.7183$, the base of the Napierian system of logarithms; $x$ the distance from the mean of $X$; and $\sigma$ the standard deviation. Since

$$y_0 = \frac{N}{\sigma\sqrt{2\pi}} \quad \text{or} \quad \frac{N}{2.5066\sigma},$$

the formula may also be written

$$y = \frac{N}{2.5066\sigma}\, e^{\frac{-x^2}{2\sigma^2}}$$

There are several methods of constructing normal curves. One is to plot the expansion of the binomial with terms equal to a fairly high power, at least the sixth, and draw a smooth curve through the points so obtained. A curve so constructed will not be exactly normal, but nearly enough so for most purposes.

The best and easiest method is to make use of the data found in tables of the curve, such as those in the Appendix. The first step is to decide on the dimensions of the graph, that is, how high it is to be at the center and how wide at the base. For example, we may wish to make it 6 units high and 12 units

wide. Therefore the two scales, vertical and horizontal, are laid off with these dimensions. Since, as has already been stated, the curve extends horizontally to infinity, a decision must be made as to how far to extend it in the graph. For one of the size here employed, a distance of 3 standard deviations, or $3\sigma$, in each direction from the center or maximum ordinate is far enough; hence twice that, or $6\sigma$, is taken as equal to 12 units, with the result that $1\sigma$ equals 2 units.

After these decisions have been made and the scales plotted as shown in Figure 12, the next step is to locate points through

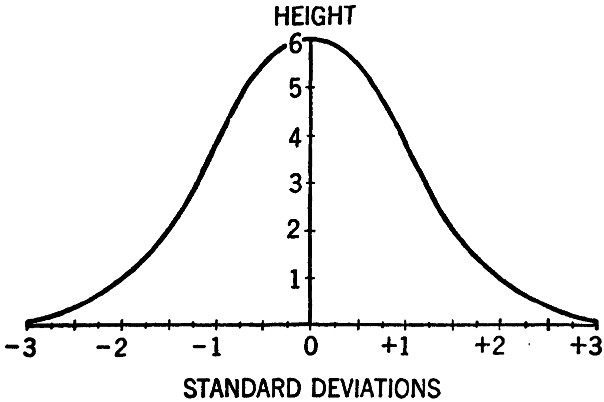



**Fig. 12. Normal Frequency Curve Constructed by Use of Data in Appendix**

which the curve may be drawn. One such point will be at the top of the maximum ordinate, or 6 units above the center of the base line. Others are found by taking a series of base-line distances from the maximum ordinate and finding the height corresponding to each. For rough work, distances at $.5\sigma$ interval are sufficient; for more exact work, they should be closer together. In this instance an interval of $.25\sigma$ will be used. Thus the first distance from $y_0$ to be used is $.25\sigma$. It is located in the first column of the standard deviation table in the Appendix and the corresponding entry in the second or height column found. This is .9692; therefore at a distance of $.25\sigma$, which here equals .5 unit, on each side of the center a point is

placed .9692 as high as the maximum ordinate or .9692 × 6., which equals 5.82, units high. For $.5\sigma$ the corresponding height entry is .8825; hence at $.5\sigma$ or 1.0 unit from the center in each direction will be a point at a height of .8825 × 6., or 5.30, units. Similarly at $.75\sigma$ or 1.5 units distance will be points .7548 × 6., or 4.53, units high; at $1.00\sigma$ or 2.0 units distance will be points .6065 × 6., or 3.64, units high; and so on until at $3.0\sigma$ or 6.0 units distance will be points .0111 × 6., or .07, units high. A smooth curve is then passed through these points, with the result shown in Figure 12.

Probably most normal curves which readers have occasion to construct will not be, as here, merely abstract, but will be those most nearly fitting actual sets of data. The construction of such best-fitting normal curves will be dealt with later, in Chapter XIII.

## Other symmetrical curves

All curves of which the halves to the right and left of the center are alike are called *symmetrical* or *isokurtic*. Symmetrical curves which also have the general characteristic of being highest at the center and decreasing in height toward the extremes are grouped into three classes on the basis of the degree of bunching or grouping around the center. This trait

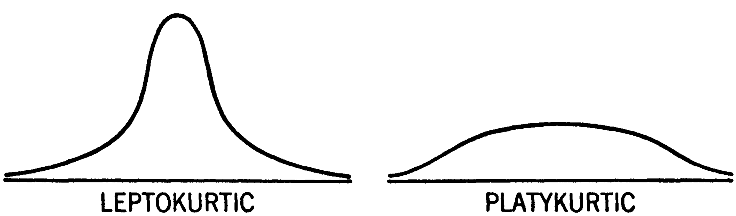



**Fig. 13. Leptokurtic and Platykurtic Curves**

is called *kurtosis*. The normal curve, in this connection termed *mesokurtic*, is considered to have average kurtosis or bunching. Curves that are flatter or less bunched near the center are known as *platykurtic*, those steeper or more bunched as *leptokurtic*. Figure 13 illustrates these two types.

Because of the variation in appearance due to choice of

different scales, moderately platykurtic and leptokurtic curves are often not distinguishable from mesokurtic ones by mere visual inspection. A formula for identifying them will be given later, in Chapter VI.

A rare type of symmetrical curve, more often encountered elsewhere than in educational work, is the *U-curve,* so called because its shape resembles that letter. In other words, it is highest at the extremes, lowest at the center. An ideal form of such a curve is shown in Figure 14. The fact that data assume

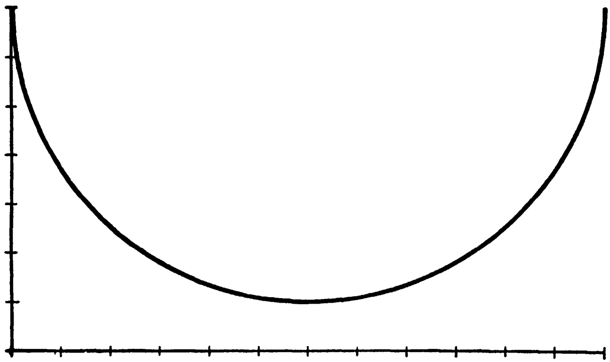


Fig. 14. U-Curve

this shape often indicates that they are heterogeneous, composed of two distinctly different types. For example, the mental test scores of honor-roll and failing pupils would, if thrown together, probably yield such a distribution.

Still another variety of symmetrical distribution is the rectangular. In it the frequencies in the various classes are the same, so that the upper line bounding the graph is straight and horizontal and, with the ends and base, forms a rectangle.

## Nonsymmetrical curves

Curves of which the two halves, right and left, are not alike are called *asymmetrical, allokurtic,* or sometimes *skew.* The latter term, also written *skewed,* is better reserved for those asymmetrical curves that may be thought of as formed by pulling a normal curve out in one direction and pushing it in on the other side. Such a curve is said to be skewed in the direc-

tion in which it extends farther from its highest point. Therefore, as shown in Figure 15, a curve extending out to the right and correspondingly bunched toward the left is positively skewed, and one extending farther to the left and bunched toward the right is negatively skewed.

A curve that is extremely skewed, so that its highest point is at or near one extreme and its lowest at the other, is sometimes

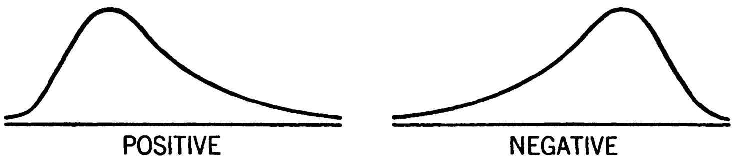



Fig. 15. Skew Curves

called a *J-curve,* from its resemblance to that letter. It may be either positively or negatively skewed.

Moderately skewed curves are frequently encountered in dealing with educational data, and extreme ones occasionally. Tests that are too hard or too easy for the average ability of those tested usually yield them, as also do various other situations. J-curves are likely to result when some factor arbitrarily cuts off the measures at one end or the other.

## Curve fitting

Especially in connection with finding the type of curve which best fits actual data, other varieties than those mentioned are often employed. A number of schemes of grouping or classifying curves have been proposed. For example, Karl Pearson, the famous English statistician who for a quarter of a century before his death a few years ago was generally ranked as the world's leader in this area, suggested a seven-fold classification. The seven were the *straight line,* the *normal curve,* the *normal ogive,* the *parabola,* the *curve of organic growth,* the *curve of organic decay,* and the *Gompertz curve.* Since the ordinary worker has little occasion to do much curve fitting and since the use of most of these involves more mathematics than the writer wishes to make necessary to mastery of this

volume, only the straight line and the normal curve will be discussed in this volume. The former, as the line of regression, will be dealt with in Chapter VIII and the latter in Chapter XIII.

## EXERCISES AND PROBLEMS

**1.** Draw a histogram for each of the following distributions·

| (a) Per cent f | (b) Score f | (c) Score f | (d) Score f |
|---|---|---|---|
| 95– 2 | 750– 4 | 260– 3 | 28– 1 |
| 90– 3 | 700– 2 | 250– 8 | 26– 2 |
| 85– 5 | 650– 6 | 240–15 | 24– 0 |
| 80– 6 | 600–12 | 230–29 | 22– 3 |
| 75– 8 | 550–19 | 220–46 | 20– 6 |
| 70–11 | 500–26 | 210–82 | 18–13 |
| 65–12 | 450–17 | 200–75 | 16–16 |
| 60– 7 | 400– 5 | 190–63 | 14–14 |
| 55– 9 | 350– 0 | 180–58 | 12–11 |
| 50– 4 | 300– 2 | 170–35 | 10– 7 |
| 45– 1 | | 160–16 | 8– 5 |
| 40– 1 | | 150–12 | |
| | | 140– 9 | |
| | | 130– 2 | |

**2.** Draw a frequency polygon for each of the distributions in Exercise 1.

**3.** Draw a smooth frequency curve for each of the distributions in Exercise 1.

**4.** Draw a smooth cumulative curve, cumulated from the bottom up, for each of the distributions in Exercise 1.

**5.** Draw a smooth cumulative curve, cumulated from the top down, for each of the distributions in Exercise 1.

**6.** Prepare a histogram to show individually each score in each distribution:

(*a*) 78, 63, 85, 66, 91, 74, 80, 94, 86, 73, 79, 69, 84, 90, 75, 83, 88, 72, 57, 85, 92, 65, 96, 84, 77, 76, 86, 81.

(*b*) 38, 46, 27, 42, 44, 37, 47, 40, 32, 45, 36, 40, 41, 37, 46, 43, 35, 41, 43, 29, 36, 43, 38, 45, 34, 39, 43, 46, 49, 37, 33, 41, 44, 32, 39.

**7.** Smooth each of the distributions in Exercise 1 by threes.

**8.** Draw the smoothed histogram for Exercise 1, Part (*a*).

**9.** Draw the smoothed frequency polygons for Exercise 1, Parts (*b*) and (*c*).

**10.** Draw the smoothed smooth frequency curve for Exercise 1, Part (*d*).

11. Draw an ogive for each of the distributions in Exercise 1.

12. Insert the following percentile lines on the ogives drawn for Exercise 11:

(a) Each tenth percentile.
(b) The 5th, 10th, 25th, 50th, 75th, 90th, and 95th percentiles.
(c) The 5th, 10th, 20th, 40th, 60th, 80th, 90th, and 95th percentiles.
(d) The 10th, 30th, 50th, 70th, and 90th percentiles.

13. Prepare an ogive representing the following two sets of scores and insert the 5th, 10th, 25th, 50th, 75th, 90th, and 95th percentile lines.

| *Score* | $f_1$ | $f_2$ |
|---|---|---|
| 140– | 1 | |
| 130– | 3 | 2 |
| 120– | 6 | 5 |
| 110– | 8 | 9 |
| 100– | 12 | 13 |
| 90– | 15 | 11 |
| 80– | 10 | 8 |
| 70– | 7 | 3 |
| 60– | 4 | 0 |
| 50– | 2 | 3 |

14. Construct a normal curve with a maximum height of 4 inches and a base from $+3\sigma$ to $-3\sigma$ of 6 inches.

15. Give two examples, not mentioned in the text, of educational data likely to form a curve of each of the following types:

(a) Normal
(b) Platykurtic
(c) Leptokurtic
(d) U-shaped
(e) Rectangular
(f) Positively skewed
(g) Negatively skewed
(h) J-shaped

16. State the type of curve which the graphic representation of each distribution in Exercises 1, 6, and 13 would most nearly approach.

17. Do the same as in Exercise 16 for the distributions in Exercises 2, 3, 4, and 5 on page 30.

## REFERENCES

Fisher, R. A., *Statistical Methods for Research Workers.* Eighth Edition. Pages 64–67. Edinburgh: Oliver and Boyd, 1941.

Garrett, Henry E., *Statistics in Psychology and Education.* Second Edition. Pages 62–76, 82–89, 98–109. New York: Longmans, Green & Company, Inc., 1937.

Guilford, J. P., *Fundamental Statistics in Psychology and Education.* Pages 65–67, 76–81. New York: McGraw-Hill Book Company, Inc., 1942.

Kenney, John P., *Mathematics of Statistics.* Part I, Chaps. II, VII; pages 107–126. New York: D. Van Nostrand Company, Inc., 1939.

Lindquist, E. F., *A First Course in Statistics.* Revised Edition. Chap. IV, pages 80–85. Boston: Houghton Mifflin Company, 1942.

Mode, Elmer B., *The Elements of Statistics.* Chaps. III, X–XI; pages 333–341. New York: Prentice-Hall, Inc., 1941.

Peters, Charles C., and Van Voorhis, Walter R., *Statistical Procedures and Their Mathematical Bases.* Chap. XV; pages 279–304. New York: McGraw-Hill Book Company, Inc., 1940.

Walker, Helen M., *Elementary Statistical Methods.* Pages 36–49. New York: Henry Holt & Company, Inc., 1943.

# CHAPTER IV

# Measures of Central Tendency

## Scope and function

The one item of information that best describes or summarizes a frequency distribution is usually a statement of the scale value about which the cases composing it tend to cluster—in other words, a statement of its central point. Such a statement of a central point is called a *measure of central tendency*, or an *average*. Although some mathematicians and statisticians use the latter term as is common in elementary-school arithmetic and among laymen in daily life—that is, as the sum of a number of cases divided by their number—it is better usage to employ it in the more general sense stated in the previous sentence. Thus averages or measures of central tendency include means, medians, mid-scores, modes, geometric means, harmonic means, and others. Of these, means and medians are most frequent in educational work.

It will avoid possible confusion in some situations if readers will think of most averages as points on the scale, not as values of any of the actual cases. They may happen to coincide with them, but only in the case of the empirical or crude mode and of the mid-score of an odd number of cases are they truly scores.

## The Mean

*Mean*, sometimes lengthened to *arithmetic mean*, is the statistical term for what is commonly called average. In other words, the mean, abbreviated *M*, is the sum of the items divided by their number. As such, it and the ordinary method of computing it are familiar to practically everyone who has completed elementary arithmetic. Its calculation, however, may fre-

quently be rendered much easier by the use of what is called the short method than it is by the procedure commonly employed.

The *short method of computing the mean* requires that an assumed or estimated mean be taken, the sum of the differences of the data from it found, this sum divided by the number of cases, and the result added algebraically to the *assumed mean.* The result is the *true mean.* This method involves no loss of accuracy and has the advantage of markedly reducing the sizes of the numbers to be handled. To illustrate it, two examples are given below. The first compares the computation of the mean of a series of ungrouped data by the short method with its computation by the long or ordinary method; the second does the same for a grouped distribution. The advantage of the short method is greater in the second case than in the first.

TABLE IV

COMPUTATION OF THE MEAN OF AN UNGROUPED SERIES BY BOTH LONG AND SHORT METHODS

| By Long Method | By Short Method | | |
|---|---|---|---|
| *Score* | *Score* | *d* | |
| 18 | 18 | | −22 |
| 81 | 81 | +41 | |
| 52 | 52 | +12 | |
| 13 | 13 | | −27 |
| 81 | 81 | +41 | |
| 15 | 15 | | −25 |
| 15 | 15 | | −25 |
| 23 | 23 | | −17 |
| 12 | 12 | | −28 |
| 98 | 98 | +58 | |
| 14 | 14 | | −26 |
| 11 | 11 | | −29 |
| 72 | 72 | +32 | |
| 37 | 37 | | −3 |
| 85 | 85 | +45 | |
| $N = 15 \lfloor 627 = \Sigma X$ | | $+229$ | $- 202 =$ |
| $M = 41.8$ | $N = 15$ | $\lfloor +27 = \Sigma d$ | |
| | | $c = +1.8$ | |
| | | $AssM = 40.$ | |
| | | $M = 41.8$ | |

The first half of Table IV illustrates the computation of the mean of 15 ungrouped scores by the long or typical elementary-school method. The formula is

$$M = \frac{\Sigma X}{N},$$

$\Sigma$, the capital Greek sigma, being the symbol for summation[1] and $X$ that for any score. The scores are added, and their sum, 627, divided by 15, their number, gives 41.8, their mean.

The second half of the table contains the same 15 cases.[2] By inspection a round number that appears to be close to the mean is chosen. It is called the *assumed mean* or *guessed mean.* No error is caused if it is not very close to the true mean, but the figures to be dealt with are smaller the closer it is. In this instance, 40 appears to be a good value to take as the assumed mean, $AssM$. Sometimes the smallest number in the series, or a "round" number slightly smaller than it, is taken as the assumed mean. So doing avoids negative differences. In this case, the value of the smallest case, 11, is so low in comparison with the largest ones that such an assumed mean would not be worth taking. In the column headed $d$, for difference or *deviation,* the remainders found by subtracting the assumed mean from each score are entered, with proper sign. Thus $18 - 40 = -22$, the first entry in that column; $81 - 40 = +41$, so this is the second entry; the others are obtained similarly. It is not necessary, but likely to avoid confusion, to place the positive deviations in one column and the negative ones in another, as here. The $d$ column is then added. In the example, the sum of the positive deviations is $+229$; that of the negative ones, $-202$; hence that of all is $+27$. This is labelled $\Sigma d$, read "the sum of the deviations." It is divided by $N$, here 15, and the result, $+1.8$, is the *correction,* abbreviated $c$, which is the amount that the assumed mean is in error. The correction is then added to the assumed mean to give the true

[1] Sometimes $S$ rather than $\Sigma$ is used as the symbol of summation.

[2] It is more convenient to have them in order of size, but the advantage gained from such arrangement is scarcely sufficient to justify the trouble of rearranging them if they are already in some other order.

mean, here 41.8, the same as found by the other method. The formula for this method may be written

$$M = AssM + c \text{ or } AssM + \frac{\Sigma d}{N}.$$

Table V presents the two methods for a grouped distribution.

TABLE V

COMPUTATION OF THE MEAN OF A GROUPED DISTRIBUTION BY BOTH LONG AND SHORT METHODS

| By Long Method | | | | By Short Method | | | |
|---|---|---|---|---|---|---|---|
| *Score* | *f* | *midp* | *f* × *midp* | *Score* | *f* | *d* | *fd* |
| 380– | 1 | 390 | 390 | 380– | 1 | +7 | + 7 |
| 360– | 2 | 370 | 740 | 360– | 2 | +6 | +12 |
| 340– | 1 | 350 | 350 | 340– | 1 | +5 | + 5 |
| 320– | 2 | 330 | 660 | 320– | 2 | +4 | + 8 |
| 300– | 4 | 310 | 1240 | 300– | 4 | +3 | +12 |
| 280– | 0 | 290 | 0 | 280– | 0 | +2 | 0 |
| 260– | 4 | 270 | 1080 | 260– | 4 | +1 | + 4 |
| 240– | 9 | 250 | 2250 | 240– | 9 | 0 | +48 |
| 220– | 9 | 230 | 2070 | 220– | 9 | −1 | − 9 |
| 200– | 5 | 210 | 1050 | 200– | 5 | −2 | −10 |
| 180– | 7 | 190 | 1330 | 180– | 7 | −3 | −21 |
| $N$ = 44 | | | 44 \| 11160 | $N$ = 44 | | | −40 |
| | | | $M$ = 253.64 | | | 44 | + 8 = $\Sigma fd$ |
| | | | | | | | $c$ = + .182 |
| | | | | | | | $i$ = 20. |
| | | | | | | | $ci$ = + 3.64 |
| | | | | | | | $AssM$ = 250. |
| | | | | | | | $M$ = 253.64 |

In it the same data employed in previous chapters are utilized. In the long method, at the left, the mid-point of each class is found, multiplied by its frequency, the sum of these products found, and this divided by $N$ to give the mean. For example, the mid-point of the top class is 390, its frequency 1, hence the product is 390; the mid-point of the next class is 370, its frequency 2, the product 740; and so on. The sum of the products, 11,160, is divided by 44 to give 253.64, the mean.

The second half of the table contains the computation by the short method, using the formula,

$$M = AssM + ci \text{ or } AssM + \frac{\Sigma fd}{N}\, i.$$

Its first two columns are the same as those in the first half. As in the ungrouped series, an assumed mean is taken. It must be at the mid-point of some class. Here it is at 250, the mid-point of the 240-class. This class then has 0 entered after it in the $d$ column. Above it are positive entries and below it negative ones. In each direction they begin with 1 for the class next to that containing the assumed mean and increase 1 for each successive class. In other words, they show how many intervals above or below, that is, larger or smaller than, the assumed mean class each of the other classes is. In the example, the highest class is 7 intervals above, the lowest 3 below. The last or $fd$ column contains the product of each $d$ entry by the $f$ for the same class. Since this product is always 0 for the class wherein the assumed mean lies, it is customary to enter the sum of the positive $fd$'s, here $+48$, in that space. The sum of the negative $fd$'s is $-40$; hence the sum of all is $+8$. Dividing this by $N$, 44, gives a correction of $+.182$, expressed in intervals. This is multiplied by 20, the interval, and the result, $+3.64$, added to the assumed mean, 250., gives the mean, 253.64.

## The median

The median is the point on the scale on each side of which half the cases lie. It is not so common in ordinary life as the mean, but because of its usually greater ease of computation and sometimes better representation of the whole distribution it has come to be employed quite often in educational work. Probably its most frequent use is in reporting test scores.

The computation of the median, abbreviated *Md*, *Med*, or sometimes *Mdn*, is illustrated in Table VI for the same data just employed for the mean. The formula is

$$Md = l + \frac{\frac{N}{2} - S}{f}\, i.$$

Since $l$, $N$, $f$, and $i$ have been used before, $S$ is the only new symbol. It stands for *partial sum*, which will be explained

shortly. In using this formula, one should start with the first term in the numerator, $\frac{N}{2}$. Here this is $\frac{44}{2}$, which equals 22.

TABLE VI

COMPUTATION OF THE MEDIAN

| Score | $f$ | |
|---|---|---|
| 380– | 1 | |
| 360– | 2 | |
| 340– | 1 | |
| 320– | 2 | |
| 300– | 4 | |
| 280– | 0 | $Md = 240 + \dfrac{\frac{44}{2} - 21}{9} 20$ |
| 260– | 4 | |
| 240– | 9 | $= 242.22$ |
| 220– | 9 | |
| 200– | 5 | |
| 180– | 7 | |
| $N =$ | 44 | |

Next the worker should obtain the partial sum, $S$, by adding the frequencies from the bottom up until they give the largest possible sum not greater than $\frac{N}{2}$. In this case $7 + 5 + 9 = 21$, which is $S$. The denominator of the fraction, $f$, is the frequency of the next class above the highest one of which the frequency was used in getting $S$. In this instance, therefore, it is 9. The lower limit of the same class, here 240, is the $l$ of the formula. Employing these values, also $i$ as 20, the formula gives

$$Md = 240 + \frac{\frac{44}{2} - 21}{9} 20 = 242.22.$$

The median may also be found by working down from the top as well as up from the bottom, but the writer does not recommend this except as a check on the other method, for which purpose it is useful. The formula for it is

$$Md = u - \frac{\frac{N}{2} - S}{f} i.$$

This gives, for the same example,

$$Md = 260 - \frac{\frac{44}{2} - 14}{9} 20 = 242.22,$$

as before.

There are cases in which the computation of the median presents certain unusual features. Although the same formula fits such instances, it seems well to illustrate some of them. This has been done in Table VII. Part A presents the most

TABLE VII

COMPUTATION OF THE MEDIAN IN CASES PRESENTING CERTAIN UNUSUAL FEATURES

| Part A | Part B | Part C | Part D | Part E |
|---|---|---|---|---|
| $S = \frac{N}{2}$ | $S = \frac{N}{2}$ and next $f = 0$ | $S = \frac{N}{2}$ and next two or more $f$'s $= 0$ | $i$ is not uniform | $f$ of lowest class exceeds $\frac{N}{2}$ |
| *Score f* | *Score f* | *Salary f* | *Pupils f* | *Score f* |
| 50– 2 | 28– 1 | 1600– 1 | 600– 1 | 180– 1 |
| 45– 3 | 26– 6 | 1550– 3 | 500– 4 | 160– 1 |
| 40– 6 | 24–14 | 1500– 5 | 400– 3 | 140– 2 |
| 35– 7 | 22– 8 | 1450– 7 | 300– 6 | 120– 0 |
| 30–12 | 20– 0 | 1400– 2 | 200– 8 | 100– 2 |
| 25–15 | 18– 4 | 1350– 0 | 100–11 | 80– 1 |
| 20–17 | 16– 7 | 1300– 0 | 75– 9 | 60– 3 |
| 15–11 | 14–10 | 1250– 0 | 50–13 | 40– 5 |
| 10– 6 | 12– 5 | 1200– 4 | 25–18 | 20–12 |
| 5– 8 | 10– 3 | 1150– 6 | 0–14 | 0–38 |
| 0– 3 | | 1100– 5 | | |
| | | 1050– 2 | | |
| | | 1000– 1 | | |
| $N = 90$ | $N = 58$ | $N = 36$ | $N = 87$ | $N = 65$ |
| $\frac{N}{2} = 45$<br>$S = 45$<br>$Md = 25.$ | $\frac{N}{2} = 29$<br>$S = 29$<br>$Md = 20 + \frac{i}{2}$<br>$= 20 + \frac{2}{2}$<br>$= 21.$ | $\frac{N}{2} = 18$<br>$S = 18$<br>$Md = 1250 + \frac{3i}{2}$<br>$= 1250 + \frac{3 \times 50}{2}$<br>$= 1325.$ | $Md =$<br>$50 + \frac{\frac{87}{2} - 32}{13} 25$<br>$= 72.12$ | $Md =$<br>$0 + \frac{\frac{65}{2} - 0}{38} 20$<br>$= 17.11$ |

frequent of these cases, that in which $S$ is just equal to $\frac{N}{2}$ and

there is no zero frequency in the next class. The formula can be applied in full; thus,

$$Md = 25 + \frac{\frac{90}{2} - 45}{15} 5 = 25,$$

but it is unnecessary to do so. Instead, as soon as it is evident that $S = \frac{N}{2}$, one may at once take the lower limit, here 25, of the next class above the highest one added in getting $S$, as the median.

Parts B and C present examples in which $S = \frac{N}{2}$ and the next class or classes above the highest one used in getting $S$ have zero frequencies. In such cases the median is equal to the lower limit of the next class above the highest one added plus one-half of the distance covered by the zero frequencies. Thus in Part B the next lower limit is 20 and one frequency of 0 occurs. Since the interval is 2, the median equals $20 + \frac{2}{2} = 21$. In Part C there are three frequencies of 0, with an interval of 50; hence $\frac{3 \times 50}{2} = 75$ should be added to 1,250, the next lower limit, to give a median of 1,325.

In Part D the interval is not uniform. The only special point to be watched here is that, for $i$ in the formula, the value of the interval within which the median falls be substituted. Hence there,

$$Md = 50 + \frac{\frac{87}{2} - 32}{13} 25 = 72.12.$$

Part E differs from most cases in that the frequency of the lowest class is greater than $\frac{N}{2}$. In such cases $l$ of the formula is the lower limit of the bottom class and $S$ is 0. In the example,

$$Md = 0 + \frac{\frac{65}{2} - 0}{38} 20 = 17.11.$$

### The mid-score or mid-measure

Many workers in this field employ the term *median* regardless of whether the data in question compose a simple or a grouped distribution. Others, however, prefer to limit its use to the latter case. In other words, they employ the term only for a measure obtained from a grouped series by the use of the formula employed in the preceding section and use *mid-score* or *mid-measure*, instead of median, for a simple or ungrouped series. This practice appears to be preferable.

The mid-score or mid-measure may be simply defined as the middle one of a series of scores or other measures arranged in order of size. If the total number is odd, it is the $\frac{N+1}{2}$th case from either end; if the total number is even, it is the mean of the two mid-most cases, that is, of the $\frac{N}{2}$th and the $\left(\frac{N}{2}+1\right)$th cases. For example, if 17 scores are arranged in order of size, the mid-score is the $\frac{17+1}{2}$th or the 9th; if 25 cases, the 13th; if 28 cases, the mean of the 14th and 15th; and so on. In dealing with pupils' papers it is frequently convenient to find it by arranging the papers in order and counting in to the middle one without making a list of the scores.

Since the mid-score is found for an ungrouped distribution and the median for a grouped one, the mid-score is usually found for a distribution containing few cases—perhaps not over 30 or 40—and the median for one with more.

### The mode

The *mode*, abbreviated *Mo*, is, as its name implies, that which is most popular or frequent. In other words, it is the point on the scale at which there are most cases. Therefore the mode of a graphed distribution is the point or scale value directly below the highest point of the curve.

The mode obtained by mere inspection, as just suggested, is called the *crude, empirical, apparent,* or *inspectional mode.* It

is almost instantaneously obtainable and frequently serves as a good typical value of the thing being measured.

In the case of tabulated distributions the crude mode cannot be exactly determined. It is customary either to state that it falls in a specified class or that it is the mid-point of that class. Thus the crude mode of the distribution given at the right would be said to be in the 50-class, or at 55. The former is preferable.

| Score | $f$ |
|---|---|
| 90– | 2 |
| 80– | 4 |
| 70– | 5 |
| 60– | 8 |
| 50– | 13 |
| 40– | 11 |
| 30– | 9 |
| 20– | 7 |
| 10– | 1 |
| $N =$ | 60 |

The *true mode* is the point at which the greatest frequency would occur if there were no grouping of measures, if all were perfectly reliable, and if no sampling were involved. It cannot actually be determined, but an approximation to it, known as the *estimated, computed,* or *refined mode,* is often found. Various formulas for this purpose have been suggested. Two of those commonly used will be given here. More exact ones exist, but they are longer and more difficult.

The most frequently used formula is $Mo = 3Md - 2M$. For distributions not very far from normal in their shape, this generally yields a close approximation to the true mode. For the distribution just given,

$$Md = 51.54 \text{ and } M = 52.33;$$

therefore,

$$Mo = 3 \times 51.54 - 2 \times 52.33 = 49.96.$$

For the distribution of 44 cases so often used,

$$Mo = 3 \times 242.22 - 2 \times 253.64 = 219.38.$$

Since, however, this distribution is not nearly normal, this value is probably not a close approximation to the true mode. The reader will notice that this formula yields a value for the mode that is on the same side of the mean as the median, but three times as far away. Sometimes, indeed, the formula is written

$$Mo = M - 3(M - Md),$$

but it is simpler to multiply, remove the parentheses, and combine into the form given above.

For quite asymmetrical distributions, the second formula for the mode is generally preferable to that given above. It is

$$Mo = l + \frac{f_a \cdot i}{f_a + f_b},$$

in which $l$ is the lower limit of the modal class, the subscript $a$ refers to the class just above the modal class and $b$ to that just below it. For the distribution on the previous page this gives

$$Mo = 50 + \frac{8 \times 10}{8 + 11} = 54.21.$$

In applying this formula to the distribution of 44 cases there is the difficulty that, since two classes each have the largest frequency, there is no one modal class. In such a case, probably the best procedure is simply to take the point that divides the two classes as the mode. Here it is 240. If the two classes with the highest frequency are not adjacent, this cannot be followed, nor can it if there are more than two classes with the same largest frequency. In such cases a single mode for the distribution has so little meaning that it is scarcely worth determining.

In cases with two or more tied frequencies as the largest, or even with two or more unequal frequencies greater than the adjacent ones, a distribution or curve may be referred to as *bimodal* if there are two, *multimodal* if there are more than two. Thus the much-used distribution of 44 cases is multimodal, but with two *major modes* and several *minor modes*. The one employed only in this section, with a single high point and no other frequency greater than both adjacent ones, is *unimodal*.

## The geometric mean

The *geometric mean* is the $N$th root of the product of $N$ measures. In formula form,

$$G, GM, \text{ or } M_G = \sqrt[N]{X_1 \cdot X_2 \cdot X_3 \cdot \ \ldots\ldots\ X_N}.$$

It is frequently employed in economics and elsewhere in connection with index numbers, but its only use in education that seems worth mentioning here is to determine rates of increase of population, enrollment, and so forth.

Its use for the purpose just mentioned ordinarily occurs when predicted increase is desired and it is assumed that the rate of increase will remain the same as it has been for the period covered by the data at hand. Since this assumption is often not valid, the predictive value of the geometric mean is quite limited. All the data necessary to find it are the figures on whatever is involved at the beginning and end of a known period. In such situations a variation of the formula given above is most convenient. It is

$$G = \sqrt[N]{\frac{f_N}{f_0}},$$

in which $N$ is the number of time units, $f_N$ is the frequency or number at the end of $N$ time units, and $f_0$ that at the beginning.

The application of the formula just given may be illustrated by assuming that the annual rate of increase in population is desired for a community which had 2,872 inhabitants in 1930 and 3,628 in 1940. Since there are ten years involved, $N = 10$, and the formula gives

$$G = \sqrt[10]{\frac{3628}{2873}} = \sqrt[10]{1.2632} = 1.0236.^{3}$$

This means that, on the average, each year's population during 1930–1940 was 1.0236 times that of the year before. In other words, the average annual increase was .0236 or 2.36 per cent.

As a second example, we may suppose that a pupil was 45 inches tall in September and 49 inches tall in June, 9 months later, and his average monthly rate of increase is desired.

$$G = \sqrt[9]{\frac{49}{45}} = \sqrt[3]{\sqrt[3]{1.0889}} = 1.0086,$$

[3] The value of $\sqrt[10]{1.2632}$ was found by the use of logarithms. Without knowledge as to how to use them, readers will find it very difficult to secure geometric means except when $N$ equals some power or product of 2 and 3. In such cases a succession of square and cube roots, as needed, can be extracted. For example, the fourth root is the square root of the square root; the sixth root is the square root of the cube root; and so on. Since this is the only procedure dealt with in this volume where the use of logarithms is likely to be necessary, and it does not actually occur very often in educational practice, their nature and use will not be explained here.

whence the average monthly rate of increase is .0086 or .86 per cent.

The same procedure may be applied to rates of decrease as well as to those of increase. In such instances $G$ comes out less than 1.00 and the difference is the rate of decrease.

## The harmonic mean

It appears in place to mention this measure, but not to discuss it at length. The *harmonic mean* is the reciprocal of the mean of the reciprocals of the measures. In formula form,

$$H,\ HM, \text{ or } M_H = \frac{N}{\Sigma\left(\frac{1}{X}\right)}.$$

It is rarely employed. For data involving both time and work units, both it and the arithmetic mean are appropriate but not directly comparable, since one is based on one unit, the other on the other unit. If one wishes an average time rate and is employing work units, he must use either the arithmetic mean of the absolute times or the harmonic mean of the time rates; if he is employing time units, he must use either the arithmetic mean of the time rates or the harmonic mean of the absolute times.

## The use of averages

Mathematical statisticians enumerate six or more criteria of a good average and discuss the various measures of central tendency in terms thereof. For present purposes a less thorough treatment seems sufficient.

The mean is, in most instances, the most reliable average. As distributions become more and more skew, however, it loses this advantage until for very skew ones it may be less reliable than the median, or even than the mode. It is likely to be more affected by one or a few large errors than is either the median or the mode. Moreover, the more heterogeneous the group of data, the less satisfactory is the mean as a representative measure. Its computation is more difficult than that of the median, but its determination is a necessary step in com-

puting several other measures that may be wanted, whereas that of the median rarely is.

The median has come to be almost universally employed as a measure of central tendency for distributions of test scores and many other kinds of data. It has the advantage of being consistent in type with the other point measures to be presented in the next chapter. If signs are neglected, the sum of deviations of scores from it is a minimum. When exact values of extreme cases are unknown it can be found, whereas the mean cannot. Also, when the units of measurement employed are not known to be equal, it is preferable. It does not assume such equality, whereas the mean does.

The mode is rarely employed in educational work except as a quick, "rough-and-ready" measure. In some instances, especially where the greatest frequency constitutes a large fraction of the total number of cases, the mode is the most typical and representative measure of central tendency. For example, if a high-school principal receives $2,700 and each of eight teachers $1,620, the mode, $1,620, is for most purposes more representative of the typical salary in that high school than is the mean, $1,740. The mid-score, also $1,620, is of course equally typical. If, however, the figures were to be employed in such a connection that the total amount of salary would be wanted, the mean would be more useful, since the total could be determined from it but not from either of the other two.

## EXERCISES AND PROBLEMS

**1.** Find the mean, by the short method, of each of the following distributions:

| (*a*) | | (*b*) | | (*c*) | | (*d*) | | (*e*) | |
|---|---|---|---|---|---|---|---|---|---|
| *Age* | *f* | *Score* | *f* | *Score* | *f* | *Score* | *f* | *Words* | *f* |
| 20–0– | 2 | 66– | 4 | 120– | 2 | 900– | 6 | 48– | 3 |
| 19–6– | 1 | 63– | 5 | 110– | 3 | 850– | 3 | 46– | 4 |
| 19–0– | 3 | 60– | 3 | 100– | 6 | 800– | 7 | 44– | 7 |
| 18–6– | 8 | 57– | 7 | 90– | 11 | 750– | 8 | 42– | 3 |
| 18–0– | 11 | 54– | 11 | 80– | 18 | 700– | 4 | 40– | 6 |
| 17–6– | 9 | 51– | 19 | 70– | 25 | 650– | 10 | 38– | 2 |
| 17–0– | 12 | 48– | 12 | 60– | 17 | 600– | 8 | 36– | 2 |
| 16–6– | 13 | 45– | 13 | 50– | 10 | 550– | 1 | 34– | 0 |
| 16–0– | 15 | 42– | 6 | 40– | 5 | 500– | 3 | 32– | 1 |
| 15–6– | 14 | 39– | 2 | 30– | 2 | 450– | 2 | 30– | 0 |
| 15–0– | 17 | 36– | 4 | 20– | 2 | 400– | 1 | 28– | 0 |
| 14–6– | 15 | 33– | 1 | 10– | 0 | | | 26– | 1 |
| 14–0– | 9 | | | 0– | 3 | | | 24– | 0 |
| 13–6– | 4 | | | | | | | 22– | 1 |
| 13–0– | 1 | | | | | | | 20– | 1 |

**2.** Find the median of each of the distributions in Exercise 1.

**3.** Find the median of each of the following distributions:

| (*a*) | | (*b*) | | (*c*) | | (*d*) | | (*e*) | |
|---|---|---|---|---|---|---|---|---|---|
| *Score* | *f* | *Score* | *f* | *Score* | *f* | *Score* | *f* | *Pupils* | *f* |
| 260– | 2 | 52– | 3 | 15– | 3 | 250– | 1 | 3500– | 1 |
| 240– | 3 | 50– | 5 | 14– | 4 | 225– | 3 | 3000– | 0 |
| 220– | 6 | 48– | 6 | 13– | 2 | 200– | 1 | 2500– | 2 |
| 200– | 8 | 46– | 11 | 12– | 6 | 175– | 4 | 2000– | 3 |
| 180– | 13 | 44– | 7 | 11– | 4 | 150– | 2 | 1500– | 4 |
| 160– | 16 | 42– | 9 | 10– | 3 | 125– | 6 | 1000– | 7 |
| 140– | 10 | 40– | 0 | 9– | 0 | 100– | 7 | 500– | 22 |
| 120– | 9 | 38– | 2 | 8– | 0 | 75– | 9 | 400– | 13 |
| 100– | 9 | 36– | 8 | 7– | 1 | 50– | 16 | 300– | 17 |
| 80– | 8 | 34– | 18 | 6– | 5 | 25– | 25 | 200– | 33 |
| 60– | 5 | 32– | 6 | 5– | 4 | 0– | 34 | 100– | 54 |
| 40– | 4 | 30– | 7 | 4– | 3 | | | 0– | 68 |
| 20– | 2 | | | 3– | 4 | | | | |
| 0– | 1 | | | 2– | 2 | | | | |
| | | | | 1– | 3 | | | | |

**4.** Find the mode, by the formula based on the mean and the median, for each of the distributions in Exercise 1.

**5.** Find the mode, by the formula based on the frequencies above and below the modal class, for each of the distributions in Exercise 1.

**6.** Find the mid-score or mid-measure of each of the following sets of data:

(*a*) 25, 3, 15, 1, 9, 18, 5, 16, 17, 11, 27, 16, 23, 19, 8, 9, 11, 23, 6, 9, 28, 19, 18, 22.

(*b*) 74, 67, 89, 75, 81, 90, 78, 85, 72, 95, 88, 61, 80, 76, 72, 93, 87, 85, 74, 77, 82.

(*c*) 104, 115, 94, 97, 109, 91, 122, 86, 103, 98, 93, 111, 101, 82, 99, 105, 97, 106, 89, 114, 108, 113, 92, 96, 88, 95, 100, 106, 99, 92.

**7.** If a school enrolled 225 pupils in 1935 and 440 in 1945, what was the average annual rate of increase?

**8.** If a pupil weighs 92 pounds on September 1 and 106 pounds on May 1, what is his average monthly rate of increase?

**9.** If the per cent of failure is reduced from 12. to 7.4 in four years, what is the average annual decrease?

## REFERENCES

Garrett, Henry E., *Statistics in Psychology and Education.* Second Edition. Pages 20–30. New York: Longmans, Green & Company, Inc., 1937.

Guilford, J. P., *Fundamental Statistics in Psychology and Education.* Pages 34–44. New York: McGraw-Hill Book Company, Inc., 1942.

Kenney, John F., *Mathematics of Statistics.* Part I, Chap. III. New York: D. Van Nostrand Company, Inc., 1939.

Lindquist, E. F., *A First Course in Statistics.* Revised Edition. Chap. V. Boston: Houghton Mifflin Company, 1942.

Mode, Elmer B., *The Elements of Statistics.* Chap. V; pages 123–137, 139–140, 141–144. New York: Prentice-Hall, Inc., 1941.

Monroe, Walter S., and Engelhart, Max D., *The Scientific Study of Educational Problems.* Pages 68–74. New York: The Macmillan Company, 1926.

Peters, Charles C., and Van Voorhis, Walter R., *Statistical Procedures and Their Mathematical Bases.* Pages 41–57. New York: McGraw-Hill Book Company, Inc., 1940.

Walker, Helen M., *Elementary Statistical Methods.* Chap. VII; pages 166–176. New York: Henry Holt & Company, Inc., 1943.

# CHAPTER V

# Other Point Measures

## General definition and function

As was suggested in the discussion of ogives in Chapter III, it is frequently useful to indicate other points in a distribution in addition to averages. Such points are based upon the division of the total number of cases in a distribution into an equal number of parts. The points which so divide it are denoted by terms formed by adding the suffix *-ile* to a stem which denotes into how many parts the distribution is divided. The single exception to this is that, as already stated in Chapter IV, the measure that divides it into two equal parts is called the median. Since to divide a distribution into any number of parts requires one less division point than there are parts, one median divides it into two parts, two tertiles into three parts, three quartiles into four parts, four quintiles into five parts, and so on. In other words, there is always one less point of any kind than its name suggests, that is, nine deciles, ninety-nine percentiles, and so on.

There is no limit to the number of such points possible, but in practice only a few are employed. Most common are the quartiles; next are the percentiles; others are relatively rare but occasionally useful.

Each set of point measures is numbered from first up to the highest. Each measure so numbered indicates the fraction of cases below the point indicated. Thus one-fourth of the cases in a distribution are below its first quartile; two-fourths are below its second quartile, which is the same as its median; three-fourths are below its third or upper quartile. Similarly, to give several examples, at random, one-tenth of the cases

lie below the first decile; seven per cent lie below the seventh percentile; two-thirds lie below the second tertile; ninety-eight per cent lie below the ninety-eighth percentile; and so on. By subtracting the fraction below any point from unity, the percent above it is easily found. Thus three-fourths are above the first quartile; one-tenth are above the ninth decile; two per cent are above the ninety-eighth percentile; and so forth.

It is frequently possible to call a point measure by any one of several names. Thus the median is the same as the second quartile, the third sextile, the fifth decile, the fiftieth percentile, and all other halfway measures. Likewise the first quartile and the twenty-fifth percentile are the same; the first decile and the tenth percentile; the third quintile, the sixth decile, and the sixtieth percentile; and so on.

These measures are regularly abbreviated by their initial letters or syllables with appropriate numerical subscripts. Thus $Q_1$ is the symbol for the first quartile, $Q_3$ for the third quartile, $Quint_1$ for the first quintile, $P_5$ or $Per_5$ for the fifth percentile, and similarly for others. Readers should note that, although the usage is sometimes employed by those unfamiliar with correct terminology in this connection, the terms ending in *-ile* are not correctly applied to the parts into which a distribution is divided. For example, the part below the first quartile should be called the first quarter or fourth; the part between it and the second quartile is the second quarter or fourth; the part below the first percentile is the first per cent or hundredth; the part between the thirty-third and the thirty-fourth percentiles is the thirty-fourth per cent or hundredth; and likewise for others.

## Computation of quartiles, percentiles, and similar measures

The method of computation of quartiles, percentiles, and similar measures is similar to that for finding the median. The only difference is that the first term in the numerator, instead of being $\frac{N}{2}$, is whatever fraction of $N$ lies below the measure to be computed. Therefore the formula for the first quartile is

$$Q_1 = l + \frac{\frac{N}{4} - S}{f} i;$$

that for the third quartile is

$$Q_3 = l + \frac{\frac{3N}{4} - S}{f} i;$$

that for the fifth percentile is

$$P_5 = l + \frac{\frac{5N}{100} - S}{f} i;$$

and similarly for others.

Table VIII presents the computation of the first and third

TABLE VIII

COMPUTATION OF QUARTILES AND PERCENTILES

| Score | f |
|---|---|
| 380– | 1 |
| 360– | 2 |
| 340– | 1 |
| 320– | 2 |
| 300– | 4 |
| 280– | 0 |
| 260– | 4 |
| 240– | 9 |
| 220– | 9 |
| 200– | 5 |
| 180– | 7 |
| | $N = 44$ |

$$Q_1 = 200 + \frac{\frac{44}{4} - 7}{5} 20 = 216.00$$

$$Q_3 = 260 + \frac{\frac{3 \times 44}{4} - 30}{4} 20 = 275.00$$

$$P_5 = 180 + \frac{\frac{5 \times 44}{100} - 0}{7} 20 = 186.29$$

$$P_{10} = 180 + \frac{\frac{10 \times 44}{100} - 0}{7} 20 = 192.57$$

$$P_{70} = 260 + \frac{\frac{70 \times 44}{100} - 30}{4} 20 = 264.00$$

$$P_{90} = 320 + \frac{\frac{90 \times 44}{100} - 38}{2} 20 = 336.00$$

quartiles and four percentiles for the series of data previously employed. The first step in determining the first quartile is to take one-fourth of $N$, here 44, which in this case is 11. $S$ is 7,

since adding the next frequency, 5, would give a sum larger than 11. The frequency of the next class is, as just stated, 5; its lower limit is 200; and $i$ is 20. Therefore,

$$Q_1 = 200 + \frac{\frac{44}{4} - 7}{5} 20 = 216.00$$

The others are found in like fashion, the first term in the numerator of each being the fraction of $N$ indicated by the name of the measure being found.

## Overlapping

Sometimes we wish to compare distributions by determining how much they overlap. To do so, we find what fraction of the cases in one overlap or extend beyond some point in the other. Overlapping may be computed in either direction, above or below, and may be past any point. The point most commonly employed for this purpose is the median; next to it are the quartiles. The fraction is usually expressed in per cent.

The data at the side may be employed to illustrate the determination of overlapping. They represent the mental ages of pupils in a fifth and a sixth grade in the same school. If we desire to find out what per cent of those in the fifth grade exceed the median of the sixth, the first step is to find this median. It is

| *MA* | *V* | *VI* |
|---|---|---|
| 14– | | 1 |
| 13– | 2 | 4 |
| 12– | 3 | 10 |
| 11– | 11 | 18 |
| 10– | 16 | 2 |
| 9– | 6 | 1 |
| 8– | 2 | |
| $N =$ | 40 | 36 |

$$11 + \frac{\frac{36}{2} - 3}{18} 1 = 11^5/_6.$$

Those of the fifth-grade pupils with mental ages above this are the 2 in the 13-class, the 3 in the 12-class, and $\frac{1}{6}$ of the 11 in the 11-class. This fraction is taken because $\frac{1}{6}$ of the interval is above $11^5/_6$, the point past which overlapping is to be measured.

$$2 + 3 + \frac{11}{6} = 6^5/_6,$$

the number of fifth graders above the sixth-grade median. This divided by 40 gives about 17 per cent.

For a second example, we may assume that the per cent of sixth graders below the first quartile of the fifth is wanted.

$$Q_1 = 10 + \frac{\frac{40}{4} - 8}{16} 1 = 10.125.$$

The number of sixth graders below 10.125 is $1 + .125 \times 2 = 1.25$, which is approximately $3\frac{1}{2}$ per cent of 36.

Another, and rarer, method of measuring overlapping is to take the difference between similar measures of central tendency of the two distributions. For the data above,

$$Md_5 = 10.75 \text{ and } Md_6 = 11.83,$$

hence the difference is 1.08. To render this relative rather than absolute, it is divided by an appropriate measure of variability of the distribution the overlapping of which is being found. Thus, if that of the sixth on the fifth is desired, 1.08 may be divided by the quartile deviation of the sixth-grade scores, the computation of which is presented in the next chapter, .83, the result being 1.30.

Bi-serial $r$, the meaning and calculation of which will be given in Chapter XII, may also be employed to measure overlapping. It appears to merit more use for this purpose than it has received. The greater its value, the less the overlapping.

## Finding percentile ranks of scores in grouped series

There are two somewhat different situations in which percentile ranks of scores in grouped distributions may be desired. Both will be explained. One is that in which a percentile rank is to be assigned to all the cases in a class, considering them as being located at its mid-point. This is very easily done. All that is necessary is to find the sum of the frequencies in all classes below the one in question plus half of its frequency, divide this by the total number of cases, and multiply by 100. In formula form, this is

$$100 \frac{S + \frac{f}{2}}{N}.$$

For example, the percentile rank of the mid-point of the 220-class in Table VIII is

$$100\,\frac{7+5+\frac{9}{2}}{44} = \frac{1650}{44} = 37.5;$$

that of the cases in the 300-class is

$$100\,\frac{7+5+9+9+4+0+\frac{4}{2}}{44} = \frac{3600}{44} = 81.82;$$

and similarly for others.

The other situation is that in which the percentile rank of a given score is wanted. The general procedure is the same as in the other case, the only difference being in the last term of the numerator. In this instance this term is found by multiplying the frequency of the class wherein the score lies by a fraction the numerator of which is the difference between the score and the lower limit and the denominator of which is the class interval. Thus the fraction represents the proportion of the cases in the class that lie below the score in question. To illustrate the procedure, a score of 217 may be used. Its percentile rank is

$$100\,\frac{7+\frac{217-200}{20}5}{44} = \frac{1125}{44} = 25.57.$$

As another example, the percentile rank or *PR* of 365 is

$$100\,\frac{7+5+9+9+4+0+4+2+1+\frac{365-360}{20}2}{44} = \frac{4150}{44} = 94.32.$$

The formula for this may be written

$$PR = 100\,\frac{S+\frac{X-l}{i}f}{N},$$

in which $X$ represents the score to be expressed as a percentile.

## Changing ranks into percentiles

It is often convenient to compare or combine the ranks of an individual in several groups, or perhaps those of different individuals in different groups. If the numbers of cases in the groups are the same, such comparisons or combinations can be made directly and simply by the use of ordinary ranks. For example, if a pupil is third in a group of 35 and seventh in another group of 35, it is at once evident that he is 4 ranks better in one than in the other and that his average rank in the two is fifth. If, however, the numbers of cases in the groups concerned differ, it is not readily apparent how much his rank in one differs from his rank in another, nor what his average rank is.

To make comparisons and combinations in such cases all ranks must be changed to the same basis. This may be any system of dividing a distribution into parts with equal numbers of cases, such as by quartiles, deciles, and so on. In practice, percentile ranks or scores are almost always used. The general formula for changing ordinary rank[1] into any other rank is

$$XR = \frac{R - \frac{1}{2}}{N} X,$$

which becomes

$$\frac{2R - 1}{2N} X$$

or

$$\frac{X \cdot R - \frac{X}{2}}{N}.$$

$R$ is the symbol for rank, when used with $X$ referring to the new rank and when alone to the ordinary rank with 1 as lowest and $N$ as highest, and $X$ is the order of the new rank. For percentile rank the formula becomes

$$PR = \frac{R - \frac{1}{2}}{N} 100 \text{ or } \frac{2R - 1}{2N} 100 \text{ or } \frac{100R - 50}{N}.$$

[1] Readers who have difficulty in assigning ordinary ranks should consult the section thereon near the first part of Chapter X.

The use of this formula may be illustrated by the case of a pupil who stands twelfth from the bottom of an algebra class of 28 members, tenth from the bottom of a Latin class of 27, sixth from the bottom of an English class of 30, and eighteenth from the bottom of a history class of 24. His best rank is evidently in history and his worst in English, but neither the differences among his ranks nor his average rank can be determined directly from the facts given. His percentile ranks may be found as follows:

$$PR_{\text{Alg}} = \frac{100 \times 12 - 50}{28} = 41.07 \quad PR_{\text{Lat}} = \frac{100 \times 10 - 50}{27} = 35.19$$

$$PR_{\text{Eng}} = \frac{100 \times 6 - 50}{30} = 18.33 \quad PR_{\text{Hist}} = \frac{100 \times 18 - 50}{24} = 72.92.$$

With these percentile ranks available, the average rank of this pupil can readily be found by averaging them, as follows:

$$\frac{41.07 + 35.19 + 18.33 + 72.92}{4} = 41.88.$$

It is sometimes desirable to be able to change percentile or other similar ranks into ordinary ranks. The general formula is

$$R = \frac{N \cdot XR}{X} + \frac{1}{2} \text{ or } = \frac{N \cdot XR + \frac{X}{2}}{X}.$$

For changing percentile ranks into ordinary ranks it becomes

$$R = \frac{N \cdot PR}{100} + \frac{1}{2} \text{ or } = \frac{N \cdot PR + 50}{100}.$$

For example, if a pupil has a percentile rank of 15 in a group of 30, his ordinary rank is

$$\frac{30 \times 15 + 50}{100} = 5$$

from the bottom.

Since a percentile or any other rank of the *-ile* type indicates the fraction of cases below it, there can be no actual case with a rank of either 0 or the whole number of divisions formed by the *-ile* points. A 0 rank would indicate a rank above no cases and below all; the latter one, above all and below none. A

score cannot, however, be either below or above itself; hence the lowest rank is that which indicates a position below all the other, or $N-1$, scores and the highest rank that which indicates a position above all the other, or $N-1$, scores. To dispose of the score for which the rank is being sought, it is considered half above and half below itself. In other words, the assumed distribution of cases is the same as that of cases in a class of a grouped distribution, as explained on p. 25. Therefore the lowest possible *-ile* rank of a case is $\frac{X}{2N}$ and the highest $X - \frac{X}{2N}$. For percentiles these become

$$\frac{50}{N} \text{ and } 100 - \frac{50}{N}.$$

Therefore the percentile rank of the lowest of 10 cases is

$$\frac{100}{2N} \text{ or } \frac{50}{10} = 5,$$

and that of the highest is

$$100 - \frac{50}{10} = 95;$$

that of the lowest of 29 cases is

$$\frac{50}{29} = 1.72,$$

and that of the highest

$$100 - \frac{50}{29} = 98.28;$$

and likewise with others.

## EXERCISES AND PROBLEMS

**1.** Compute the first and third quartiles, the seventh, fifteenth, fortieth, and ninetieth percentiles, of the data in each part of Exercise 1 on page 69.

**2.** Compute the first and third quartiles, the fifth, tenth, and ninety-eighth percentiles, of the data in each part of Exercise 3 on page 69.

**3.** Determine the per cent of overlapping of each distribution

above the median and also below the first quartile of the distribution paired with it:

(a)

| *IQ* | *Fast* | *Aver.* |
|---|---|---|
| 140– | 2 | |
| 130– | 4 | |
| 120– | 18 | 1 |
| 110– | 14 | 5 |
| 100– | | 28 |
| 90– | | 26 |
| 80– | | 4 |

(b)

| *Score* | *XI* | *XII* |
|---|---|---|
| 115– | | 2 |
| 110– | 1 | 4 |
| 105– | 3 | 6 |
| 100– | 7 | 12 |
| 95– | 19 | 11 |
| 90– | 24 | 9 |
| 85– | 16 | 3 |
| 80– | 8 | 1 |
| 75– | 3 | 0 |
| 70– | 2 | 1 |

4. Find the percentile rank of the mid-point of each class in each part of Exercises 1 and 3 on page 69.

5. Find the percentile rank of the following scores in the indicated parts of Exercise 1 on page 69:

(*a*) 13–10, 14–9, 16–2, 18–0, 19–1.
(*b*) 37, 42, 50, 58, 68.
(*c*) 47, 69, 76, 95, 114.

6. Find the average percentile rank of a pupil whose ordinary ranks from the bottom up are as follows: seventh among 25, third among 22, eleventh among 26, eighth among 19, fourteenth among 32.

7. How many percentile points is each of the following pupils above or below each of the other two?

(*a*) John, sixteenth in a group of 42.
(*b*) James, ninth in a group of 27.
(*c*) Joseph, tenth in a group of 30.

8. If a pupil has a percentile rank of 61.36 in a group of 22 and one of 30 in a group of 35, what is his ordinary rank in each instance?

9. What are the percentile ranks of the highest and lowest of each of the following numbers of cases? (*a*) 32; (*b*) 6; (*c*) 250.

## REFERENCES

Garrett, Henry E., *Statistics in Psychology and Education.* Second Edition. Pages 76–82. New York: Longmans, Green & Company, Inc., 1937.

Guilford, J. P., *Fundamental Statistics in Psychology and Education.* Pages 67–75. New York: McGraw-Hill Book Company, Inc., 1942.

Lindquist, E. F., *A First Course in Statistics.* Revised Edition. Chap. III. Boston: Houghton Mifflin Company, 1942.

Mode, Elmer B., *The Elements of Statistics*. Pages 137–139, 140. New York: Prentice-Hall, Inc., 1941.

Monroe, Walter S., and Engelhart, Max D., *The Scientific Study of Educational Problems*. Pages 73–74. New York: The Macmillan Company, 1926.

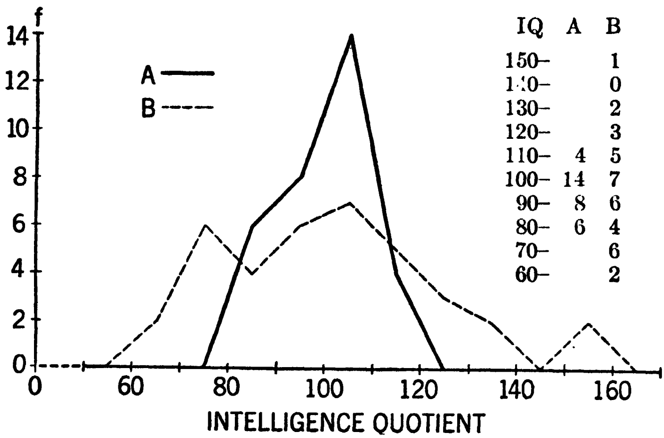
Walker, Helen M., *Elementary Statistical Methods*. Pages 59–74. New York: Henry Holt & Company, Inc., 1943.

## CHAPTER VI

# Measures of Variability

### Function and scope

Although the one most significant and descriptive item of information about a distribution is generally a measure of its central tendency, this alone is quite inadequate to portray it. Two sets of data may have the same average but differ greatly in how widely they spread about that point. This is illustrated by Figure 16 and the accompanying distributions, which show




| IQ | A | B |
|---|---|---|
| 150– | | 1 |
| 140– | | 0 |
| 130– | | 2 |
| 120– | | 3 |
| 110– | 4 | 5 |
| 100– | 14 | 7 |
| 90– | 8 | 6 |
| 80– | 6 | 4 |
| 70– | | 6 |
| 60– | | 2 |

**Fig. 16. Curves with Same Average But Different Spreads**

the distributions of intelligence quotients of two groups of pupils. Both average 100, but group A is a compact, homogeneous group of normal and near-normal pupils, whereas group B includes a range from high-class morons up to quite superior children. This difference would not be revealed by comparison of their averages.

There are other descriptive facts that might be given for any distribution, but usually the next most significant to a measure of central tendency is one of *variability*, also called *variation, dispersion, deviation, departure, discrepancy, spread, scatter,* and *fluctuation.* There are a number of measures of this characteristic. Of these the *range*, the *quartile deviation*, the *10-90 percentile range*, the *mean deviation*, the *standard deviation*, and the *median deviation* will be presented in this chapter.

The measures of variability just named are absolute numerical expressions, found in terms of the scoring system used for whatever is being measured, that express the spread or scatter of the data in terms of base-line units. As units of distance they may be employed to measure and state distance from one case to another, from a case to an average, from a case to a quartile, percentile, or other similar point, indeed from any point or value in the distribution to any other. They are frequently used as measures of distances from averages, or sometimes from other points, that include specified per cents or other fractions of the whole number of cases. Their interpretation in this connection is usually based on the assumption of normal distribution and may be in error to the extent that this assumption is not fulfilled.

In addition to absolute measures of variability there are relative measures. The *coefficient of variability* is the only one that will be given.

## The range

The *range*, or *absolute range*, is defined as the distance from the lowest case to the highest. It may, therefore, be obtained by simple subtraction. For example, the range of the eighteen salaries given on p. 19 is \$1,900 − \$1,250 = \$650.

In the case of a grouped distribution, the exact values of the two extreme cases are not known; hence the range cannot be determined so easily as for a simple or ungrouped distribution. Usually it is sufficiently close to state the range of such a distribution approximately by inspection and simple subtraction. Thus for the distribution of 44 cases used in previous

chapters, which begins with the 180-class and ends with the 380-class, the range would be stated as approximately 200.

If a more accurate range than that just suggested is desired, it should be found by determining the values of the two extreme cases according to the assumption of distribution within the classes, as explained in Chapter II. In the distribution just referred to, the lowest of the 7 cases in the 180-class is assumed to have a value of

$$180 + \frac{20}{2 \times 7} = 181.43$$

and the 1 case in the 380-class to have a value of 390; whence the range is

$$390 - 181.43 = 208.57.$$

Since a change in the value of either of the extreme cases affects the range, and may affect it largely, the range does not possess high stability. In fact, it is the most unreliable of the measures of variability mentioned in this chapter. Largely for this reason it is not very frequently employed.

## The quartile deviation

The *quartile deviation*, or *semi-interquartile range*, is one-half of the distance between the first and third quartiles. Its formula, therefore, is

$$Q, \text{ or rarely } QD, = \frac{Q_3 - Q_1}{2}.$$

Since the middle 50 per cent of the cases in a distribution lie between the first and the third quartiles, the quartile deviation is half the distance that includes the middle half of the cases. If the distribution is symmetrical, one-fourth of the cases are within $1Q$ of the median in each direction, or one-half in both. The more asymmetrical a distribution is, the more is the fraction of cases within $1Q$ of the median liable to vary from one-half.

The application of the formula may be illustrated by employing the data used on p. 73 and elsewhere. There $Q_1$ was found to be 216 and $Q_3$ to be 275; whence

$$Q = \frac{275 - 216}{2} = 29.5.$$

Since this distribution is quite asymmetrical, it is not probable that just half the cases lie within 29.5 points of the median, 242.22. Actually, despite the lack of symmetry, if the position of each case is determined according to the assumption regularly employed for grouped tabulations, it is found that almost exactly one-half the cases are between

$$242.22 - 29.5 = 212.72$$

and

$$242.22 + 29.5 = 271.72.$$

As implied by portions of the preceding paragraphs, the quartile deviation is associated with the median when with any measure of central tendency. This is because it is derived from measures, the quartile points, found by the same general method as the median.

## The 10-90 percentile range

As its name implies, the *10–90 percentile range*, symbolized by $D_{10-90}$, is the distance from the tenth to the ninetieth percentile. Therefore it includes the middle 80 per cent of the cases. Because it is based upon a larger fraction of the cases than is the quartile deviation, it is more reliable than that measure, and because it does not depend upon either of the single extreme measures, it is more reliable than the range. Despite this relatively high reliability, however, it has not come into common use. Also, it is rarely employed as a distance from an average.

The formula,

$$D_{10-90} = P_{90} - P_{10},$$

may be illustrated for the same data as were used for the quartile deviation. For them, $P_{90}$ has been found to be 336 and $P_{10}$ to be 192.57, hence

$$D_{10-90} = 336 - 192.57 = 143.43.$$

## The mean deviation

The *mean deviation, MD*—also called the *average deviation, AD*, and the *mean absolute deviation*—is the mean of the deviations, with their signs neglected, from the mean, median, or

mid-score. It is more often found about the mean than about the median or mid-score, but it is a minimum about either of the latter. In educational work the mean deviation is not used so often as the quartile, standard, and median deviations. In a normal distribution approximately 57.5 per cent of the cases are within $1MD$ of the average.

To compute the mean deviation about the mid-score of a simple series, the formula $\frac{S_2 - S_1}{N}$ may be employed. In it $S_2$ is the sum of all the cases larger than the mid-score and $S_1$ the sum of all smaller. For the ungrouped series in Table IV, on page 56, the mid-score is 23. Hence

$$MD_{Mids} = \frac{(81+52+81+98+72+37+85)-(18+13+15+15+12+14+11)}{15} = 27.2.$$

It may, of course, also be obtained by finding the deviation of each case from the mid-score, 23, summing these deviations, and dividing by $N$, here 15. In most instances, however, the former method is easier.

The mean deviation about the mean of a simple series is usually best found by a formula easily applied in connection with the short method of finding the mean. It is

$$MD_M = \frac{\Sigma|d| + c(N_b - N_a)}{N}.$$

The symbol $|\ |$ indicates that the summation is absolute, that is, with signs neglected; $N_b$ is the number of cases below the mean and $N_a$ the number above it. The correction, $c$, keeps its sign. Also, it must be small enough that no case falls between the assumed mean and the true mean. If it is not, another assumed mean must be taken. In this case, that is for the data on p. 56, $c$ is satisfactory. $\Sigma|d|$ is the sum of the positive deviations, 229, and the negative ones, 202. The number of cases below the mean, 41.8, is 9; that above it, 6. Therefore,

$$MD_M = \frac{229 + 202 + 1.8(9 - 6)}{15} = 29.09.$$

The method of computing the mean deviation of the measures

in a grouped distribution differs from that just given only in detail. The formula includes the terms in that just used with additions. It is

$$MD = \frac{\Sigma|fd| + c(N_b - N_a) + (.25 + c^2)N_m}{N}\, i.$$

The new symbol, $N_m$, refers to the number of cases or frequency in the class containing the mean or median. $N_b + N_a + N_m$ must equal $N$. The assumed mean or median must be so taken that $c$ is less than $+.5$, but not less than $-.5$. Applying this formula to the data in Table V, on page 58, we have

$$MD_M = \frac{48 + 40 + .182(21 - 14) + (.25 + .182^2)9}{44}\, 20 = 41.74$$

The method of finding the mean deviation about the median is the same. Since, however, the computation of the median does not involve the use of a correction, this must be found. It is

$$c_{Md} = \frac{Md\text{-}AssMd}{i}.$$

The assumed median must be taken at the mid-point of the same class in which the median falls, in order that the value of $c$ fall within the limits stated above. If, as is usually true, the mean and median lie in the same class, the same assumed mean and 0-deviation class is used for both, with the result that $c$ is the only term in the formula that differs. Since this is true for the data in Table V, and since for them

$$c_{Md} = \frac{242.22 - 250.}{20} = -.389,$$

the formula gives for them

$$MD_{Md} = \frac{48 + 40 - .389(21 - 14) + (.25 + .389^2)9}{44}\, 20 = 40.40.$$

This, it will be noted, is slightly less than $MD_m$, in accord with the fact that the mean deviation about the median is a minimum.

Very careful workers sometimes apply a correction to the mean deviation of a grouped distribution as obtained by the formula above in order to make it more nearly correct for the actual or ungrouped cases. This correction consists in the sub-

traction of approximately the interval divided by ten times the number of classes. If $n$ is used for the number of classes, the formula is

$$MD_{corr} = MD_{obt} - \frac{i}{10n}.$$

For the example above,

$$MD_{M_{corr}} = 41.74 - \frac{20}{10 \times 11} = 41.56$$

and

$$MD_{Md_{corr}} = 40.40 - \frac{20}{10 \times 11} = 40.22.$$

## The standard deviation

The *standard deviation,* also called the *root-mean-square deviation,* is rated by most statisticians as the most important measure of variability. It is the most reliable measure of deviation and is a quantity often needed for other statistical computations and procedures. It is associated with the mean if with any measure of central tendency.

The standard deviation is the square root of the mean of the squares of the deviations from the mean of the distribution. In a normal distribution, about 68.27 per cent of the cases fall within one standard deviation of the mean, half or 34.13 per cent on each side.

The formula for the standard deviation of a simple series is

$$\sigma\text{, or } SD, = \sqrt{\frac{\Sigma d^2}{N}} \text{ or } \sqrt{\frac{\Sigma x^2}{N}}.$$

In this form it requires that deviations from the true mean be found. Unless the true mean is a "round" number, it is easier to take an assumed mean, as in the second part of Table IV. The formula in this case becomes

$$\sigma = \sqrt{\frac{\Sigma d^2}{N} - c^2}.$$

Sometimes the actual scores rather than deviations are used, in which case it is

$$\sigma = \sqrt{\frac{\Sigma X^2}{N} - M^2}.$$

Table IX illustrates the computation of the standard devia-

TABLE IX

COMPUTATION OF THE STANDARD DEVIATION OF AN UNGROUPED SERIES BY USE OF SCORES AND OF DEVIATIONS

| By Scores | | By Deviations | | | |
|---|---|---|---|---|---|
| $X$ | $X^2$ | $X$ | $d$ | | $d^2$ |
| 18 | 324 | 18 | | −22 | 484 |
| 81 | 6561 | 81 | +41 | | 1681 |
| 52 | 2704 | 52 | +12 | | 144 |
| 13 | 169 | 13 | | −27 | 729 |
| 81 | 6561 | 81 | +41 | | 1681 |
| 15 | 225 | 15 | | −25 | 625 |
| 15 | 225 | 15 | | −25 | 625 |
| 23 | 529 | 23 | | −17 | 289 |
| 12 | 144 | 12 | | −28 | 784 |
| 98 | 9604 | 98 | +58 | | 3364 |
| 14 | 196 | 14 | | −26 | 676 |
| 11 | 121 | 11 | | −29 | 841 |
| 72 | 5184 | 72 | +32 | | 1024 |
| 37 | 1369 | 37 | | −3 | 9 |
| 85 | 7225 | 85 | +45 | | 2025 |
| $N = 15 \boxed{627} = \Sigma X$ | $N = 15 \boxed{41141} = \Sigma X^2$ | | +229 | −202 | $N = 15 \boxed{14981} = \Sigma d^2$ |
| $M = 41.8$ | 2742.73 | | $N = 15 \boxed{+27} = \Sigma d$ | | 998.73 |
| | $M^2 = 1747.24$ | | $c = +1.8$ | | $c^2 = 3.24$ |
| | $\sigma^2 = 995.49$ | | $AssM = 40.$ | | $\sigma^2 = 995.49$ |
| | $\sigma = 31.55$ | | $M = 41.8$ | | $\sigma = 31.55$ |

tion of the same simple series used in Table IV and elsewhere by both the last two formulas. In each case there are the same columns as in the table just referred to, with an additional one of squares added. The additional column in the part at the left contains the squares of the scores. Their sum, 41,141, is divided by $N$, 15, to give 2,742.73. From this 1,747.24, the square of 41.8, the mean, is subtracted, leaving 995.49. This is the square of the standard deviation; hence its square root, 31.55, is the standard deviation.

The additional column in the part at the right contains the squares of the deviations. Its sum is 14,981. Dividing this by $N$, 15, gives 998.73. The square of the correction, 1.8, is 3.24, and subtracting this from 998.73 leaves 995.49, the same value for $\sigma^2$ as was obtained previously. Therefore $\sigma$, as before, $= 31.55$.

The computation of the standard deviation of a grouped distribution is shown in Table X, which repeats the work for the computation of the mean shown in the second half of Table V on page 58 with the omission of the last steps, which are

TABLE X

COMPUTATION OF STANDARD DEVIATION OF A GROUPED DISTRIBUTION

| | $f$ | $d$ | $fd$ | $fd^2$ |
|---|---|---|---|---|
| 380– | 1 | +7 | +7 | 49 |
| 360– | 2 | +6 | +12 | 72 |
| 340– | 1 | +5 | +5 | 25 |
| 320– | 2 | +4 | +8 | 32 |
| 300– | 4 | +3 | +12 | 36 |
| 280– | 0 | +2 | 0 | 0 |
| 260– | 4 | +1 | +4 | 4 |
| 240– | 9 | 0 | +48 | 0 |
| 220– | 9 | −1 | −9 | 9 |
| 200– | 5 | −2 | −10 | 20 |
| 180– | 7 | −3 | −21 | 63 |
| | $N = 44$ | | −40 | $44 \boxed{310} = \Sigma fd^2$ |
| | | | $44 \boxed{+8} = \Sigma fd$ | $S^2 = 7.0455$ |
| | | | $c = +.182$ | $c^2 = .0331$ |
| | | | | $\sigma^2 = 7.0124$ |
| | | | | $\sigma = 2.648$ intervals |
| | | | | $i = 20.$ |
| | | | | $\sigma = 52.96$ units |

unnecessary here, and with certain additions. The first of these is the column headed $fd^2$. Each entry therein is the product of the $d$ and $fd$ entries in the same row. Thus $7 \times 7 = 49$, the first entry; $6 \times 12 = 72$, the second entry; and so on. Since each entry is the product of two like-signed numbers, all are positive. This column is totalled, giving 310. This sum is divided by $N$, 44, and the result, 7.0455, is labelled $S^2$. From this .0331, which is the square of the correction, is subtracted, leaving 7.0124. The quantity thus obtained is the square of the standard deviation, so its square root is extracted. This, 2.648, is the standard deviation in terms of intervals. Since it is regularly desirable to express it in the units employed, 2.648 is multiplied by 20, the interval, giving a value of 52.96 units.

Although in the large majority of cases values of the standard deviation found as described above are employed, there are several corrections that may be applied to produce more nearly correct values. Generally their effect is so slight in comparison with other possible sources of error that they are neglected, but it seems well to mention some of them. The most important, often called Sheppard's correction, approximately corrects the error introduced by assuming all cases to be at class mid-points. The formula is, in terms of intervals,

$$\sigma_{corr} = \sqrt{\sigma^2_{obt} - \frac{1}{12}}$$

and, in terms of units,

$$\sigma_{corr} = \sqrt{\sigma^2_{obt} - \frac{i^2}{12}}.$$

For the example in Table X this gives

$$\sigma_{corr} = \sqrt{2.648^2 - \frac{1}{12}} = 2.632 \text{ intervals,}$$

or

$$\sqrt{52.96^2 - \frac{20^2}{12}} = 52.64 \text{ units.}$$

As here, corrected $\sigma$ is always less than obtained $\sigma$.

If two series of measures of the same trait, such as two duplicate forms of a test, are available, the effect of variable errors upon the standard deviation is eliminated by using $d_1 \times d_2$, instead of $d^2$, $d_1$ being the deviation of the case in one series and $d_2$ that in the other. This can easily be applied to ungrouped data, but with some difficulty to grouped ones. Its effect is to reduce the obtained value slightly if the two series correlate closely, more if they do not.

If the coefficient of reliability, to be explained in Chapter IX, of a series of measures is known, the value of the standard deviation of true or perfectly reliable measures[1] can be estimated by the formula,

$$\sigma_{\infty} = \sigma_{obt}\sqrt{r_{\text{II}}},$$

in which $\sigma_{\infty}$ is the true standard deviation and $r_{\text{II}}$ the coefficient of reliability.

If the standard deviation of a sample is taken as an estimate of that of the whole population from which the sample was chosen—a common practice—a more accurate value is obtained if the sum of the squares is divided by $N - 1$ instead of by $N$. So doing naturally serves to increase the value of $\sigma$. Since the effect of using $N - 1$ rather than $N$ becomes less and less as the value of $N$ increases, many workers do not trouble to subtract the 1 if the number of cases is fairly large.

It is sometimes convenient to be able to find the standard deviation of the combined score on two or more duplicate forms of a test when it is known for one. To do so the coefficient of reliability must also be available. For $n$ forms of a test the formula is

$$\sigma_{1+\text{I}+\ldots n} = \frac{\sigma_1 + \sigma_{\text{I}} + \ldots \sigma_n}{\sqrt{\dfrac{n}{1 + (n-1)r_{\text{II}}}}}.$$

For only two forms this becomes

$$\sigma_{1+\text{I}} = \frac{\sigma_1 + \sigma_{\text{I}}}{\sqrt{\dfrac{2}{1 + r_{\text{II}}}}}.$$

[1] True or perfectly reliable measures cannot actually be found, but certain facts about them may be determined. They will receive more attention later.

If the standard deviations of the forms are equal, these formulas reduce to

$$\sigma_n = \sigma_1\sqrt{n + n(n-1)r_{11}}$$

and

$$\sigma_{1+1} = \sigma_1\sqrt{2(1 + r_{11})},$$

respectively. The average standard deviation of $n$ forms is

$$\sigma_1\sqrt{\frac{1}{n} + \left(1 - \frac{1}{n}\right)r_{11}}$$

and that of only two forms is

$$\sigma_1\sqrt{\frac{1}{2}(1 + r_{11})}.$$

## The variance

The *variance*, or *mean square deviation*, is the square of the standard deviation. It has no regularly employed symbol other than $\sigma^2$. It may be employed as a measure of variability and is receiving increasing use, especially in connection with the *analysis of variance* and of *covariance*, topics too broad in scope to receive adequate treatment in this volume. Also, it is utilized in the interpretation of coefficients of correlation. For the data in Table X, it is $52.96^2$, or 2,804.76.

## The median deviation

The *median deviation* (*MdD*), rarely called the *probable deviation*, is just what its name implies, the median of the deviations about the mean. Hence half the deviations from the mean exceed it and half are less; in other words, half the cases in a distribution are within 1 *MdD* of the mean and half are farther from the mean.

The approved method of finding the median deviation is not the apparently direct one of tabulating the deviations and finding their median, but is rather based upon the fact that for a normal distribution it is a fixed fraction, .6745, of the standard deviation. Applying this gives $MdD = .6745 \times 31.55 = 21.28$ for the data in Table IX and $.6745 \times 52.96 = 35.72$ for those in Table X. The formula $MdD = .6745\sigma$ is used regularly,

despite the fact that the more a distribution departs from normality, the greater the probable inaccuracy resulting from it.

For a normal distribution the quartile and median deviations are the same and for any other shape not extremely asymmetrical or irregular their values are likely to differ only slightly.

Unfortunately the term *probable error* (*PE*) has come to be very commonly applied to the median deviation. If used at all, it should be employed only when the cases dealt with are actually errors. In this sense it is explained and used later in the book.

## Comparison of measures of variability

Figure 17 presents a normal curve with the distances shown that are covered by the measures of variability so far discussed. The range and 10–90 percentile range extend the whole distance

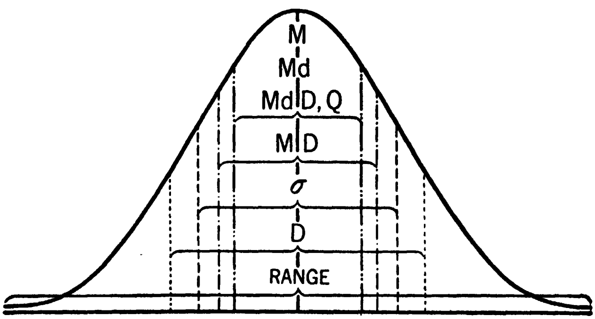



**Fig. 17. Normal Curve with Measures of Variability**

between their respective limits; whereas the others extend their respective distances in both directions from the mean or median, which, of course, coincide in a normal curve. The largest is the range, which includes all the cases; next in size is the 10–90 percentile range, which takes in the middle 80 per cent; then comes the standard deviation, which when laid off in both directions from the mean embraces 68.27 per cent; next is the mean deviation, which similarly includes 57.5 per cent about either the mean or the median; finally the median deviation includes the middle 50 per cent about the mean and the quartile deviation the same about the median.

Table XI shows the computation and relative sizes of these

TABLE XI

COMPUTATION OF MEASURES OF VARIABILITY

| | f | d | fd | fd² |
|---|---|---|---|---|
| 95– | 1 | +5 | +5 | 25 |
| 90– | 3 | +4 | +12 | 48 |
| 85– | 6 | +3 | +18 | 54 |
| 80– | 11 | +2 | +22 | 44 |
| 75– | 17 | +1 | +17 | 17 |
| 70– | 23 | 0 | +74 | 0 |
| 65– | 16 | −1 | −16 | 16 |
| 60– | 12 | −2 | −24 | 48 |
| 55– | 9 | −3 | −27 | 81 |
| 50– | 6 | −4 | −24 | 96 |
| 45– | 4 | −5 | −20 | 100 |
| 40– | 2 | −6 | −12 | 72 |
| $N =$ | 110 | | −123 | 110 ⌊601 |

$110 \,\overline{)\,-49}$

$c = -.45$

$S^2 = 5.4636$

$c^2 = .2025$

$\sigma^2 = 5.2611$

$\sigma = 2.294$ intervals

$i = 5.$

$\sigma = 11.47$ units

Range = 97.5 − 41.25 = 56.25

$Q_3 = 75 + \frac{82.5 - 72}{17}5 = 78.09 \qquad Q_1 = 60 + \frac{27.5\text{–}21}{12}5 = 62.71$

$Q = \frac{78.09 - 62.71}{2} = 7.69$

$P_{90} = 80 + \frac{99 - 89}{11}5 = 84.55 \qquad P_{10} = 50 + \frac{11 - 6}{6}5 = 54.17$

$D_{10\text{–}90} = 84.55 - 54.17 = 30.38$

$$MD_M = \frac{74 + 123 - .45(49 - 38) + (.25 + .45^2)23}{110}5 = 9.20$$

$Md = 70 + \frac{55 - 49}{23}5 = 71.30 \qquad c_{Md} = \frac{71.30 - 72.5}{5} = -.24$

$$MD_{Md} = \frac{74 + 123 - .24(49 - 38) + (.25 + .24^2)23}{110}5 = 9.16$$

$MdD = .6745 \times 11.47 = 7.74$

measures for a set of grouped data. Their calculation follows the formulas and procedures already explained. Since the distribution is skew, but not greatly so, the relationships existing among the measures of variability and their interpretation are not just those assumed on the supposition of normality, but do not differ greatly from them. For example, the quartile and median deviations are not exactly equal, as they would be for a normal distribution, but differ slightly.

Figure 18 graphically represents the data in Table XI, in histogram form, and their measures of variability. Since the

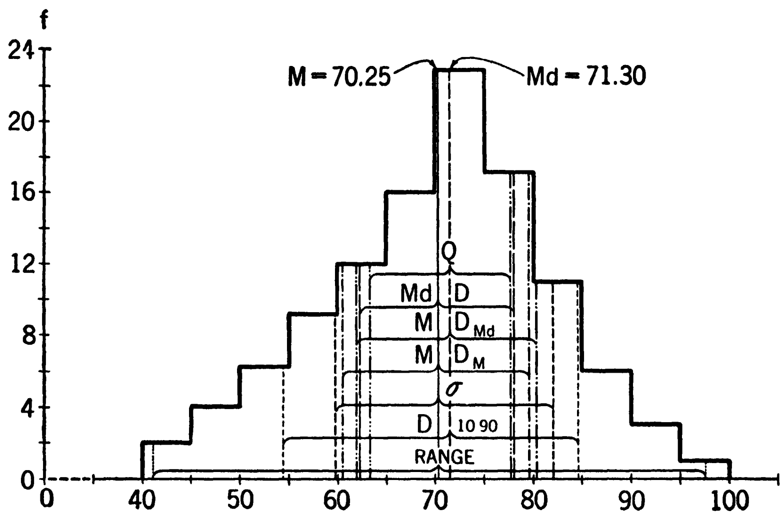



**Fig. 18. Graphic Representation of Data in Table XI, Showing Measures of Variability**

distribution is not far from normal, the areas included by the latter are approximately the same as those shown in Figure 18.

In the discussions of the various measures of variability, the per cents of the cases in a normal distribution included by distances from an average equal to them have been stated. Not only is this known for a distance equal to each such measure, but also for any distance expressed in terms of any of them except the range. The tables in the Appendix give this, among other facts, for distances expressed in terms of

standard and median or quartile deviations; but it seems in place to present here a brief table showing the per cents of cases included within distances from the average equal to even half and integral values of each of the five measures. Table XII

TABLE XII

PER CENTS OF CASES IN A NORMAL DISTRIBUTION INCLUDED WITHIN THE GIVEN DISTANCES FROM AN AVERAGE

| Number of Deviations | $MdD$ or $Q$ | $MD$ | $\sigma$ | $D_{10-90}$ |
|---|---|---|---|---|
| .5 | 26.40 | 31.01 | 38.30 | 80.00 |
| 1.0 | 50.00 | 57.51 | 68.27 | 98.96 |
| 1.5 | 68.84 | 76.86 | 86.64 | 99.99 |
| 2.0 | 82.27 | 88.95 | 95.45 | 100.00– |
| 2.5 | 90.82 | 95.39 | 98.76 | 100.00– |
| 3.0 | 95.70 | 98.33 | 99.73 | 100.00– |
| 3.5 | 98.19 | 99.48 | 99.95 | 100.00– |
| 4.0 | 99.30 | 99.86 | 99.99 | 100.00– |
| 4.5 | 99.76 | 99.97 | 100.00– | 100.00– |
| 5.0 | 99.93 | 99.99 | 100.00– | 100.00– |

gives these facts for from .5 to 5.0 of each. The entries in its second column, for example, show that 26.40 per cent of the cases in a normal distribution, or of the area under a normal curve, fall within $.5MdD$ or $Q$ of the average, 50.00 per cent of the cases within $1.0MdD$ or $Q$ of the average, 68.84 per cent within $1.5MdD$ or $Q$ of it, and so on down until 99.93 per cent are within $5.0MdD$ or $Q$ of the average. To find the per cents within the given distances of the average in one direction only, halves of these entries should be used. Thus 13.20 per cent of them are within $.5MdD$ or $Q$ of the average in one direction, and so on. The last three columns give similar per cents for other measures of variability.

It is also sometimes useful to know the numerical relationships existing among the measures of variability themselves. For a normal distribution these are as follows:

$$
\begin{aligned}
Q &= .2632D_{10-90} \text{ or } .8453MD \text{ or } .6745\sigma \text{ or } 1.0000MdD \\
D_{10-90} &= 3.8001Q \text{ or } 3.2124MD \text{ or } 2.5631\sigma \text{ or } 3.8001MdD \\
MD &= 1.1829Q \text{ or } .3113D_{10-90} \text{ or } .7979\sigma \text{ or } 1.1829MdD \\
\sigma &= 1.4826Q \text{ or } .3902D_{10-90} \text{ or } 1.2533MD \text{ or } 1.4826MdD \\
MdD &= 1.0000Q \text{ or } .2632D_{10-90} \text{ or } .8453MD \text{ or } .6745\sigma
\end{aligned}
$$

For distributions only moderately skew they do not differ greatly from these figures.

The range is not included above because, even for a normal distribution, its relationship to the other measures of variability is not constant. The reason for this lack of constancy is that a normal distribution theoretically has a range of infinity and there is no accepted basis of assuming a finite value for it that yields the same result for all normal distributions. Instead, the usual basis upon which such an assumption is made is such that the value given the range depends on the number of cases. The assumption is that the range is the distance needed to include $(N - .5)$ cases, or $\frac{N - .5}{N}$ of the area under the curve. That is, if there are 18 cases in the distribution, the range is assumed to be the distance necessary to include the middle 17.5 of them, or the middle $\frac{17.5}{18}$ of the area under the curve; if there are 45, it includes the middle 44.5 of them; and similarly for any value of $N$.

By finding the distance, in one direction from the average, that includes $\left(.5 - \frac{1}{4N}\right)$ of the area under the curve and doubling it, the assumed range for any number of cases can be found. This has been done for a few representative numbers and the results presented in Table XIII. This table shows, for example, that the assumed range of a normal distribution of 10 cases is approximately $5.8Q$, $1.5D_{10-90}$, $4.9MD$, $3.9\sigma$, or $5.8MdD$. In connection with this, as in many other situations, there are usually only small errors involved in applying the same figures and interpretations to moderately skewed distributions as are applicable to normal ones.

## The relative merits and use of measures of variability

The range is easily determined, readily understood, and may be used with any average. On the other hand, it is the least reliable and takes no account of the general shape of the distribution. On the whole, it is the least valuable measure of variability, but may often be used to supplement others.

TABLE XIII

ASSUMED RANGES OF NORMAL DISTRIBUTIONS CONTAINING VARIOUS NUMBERS OF CASES

| $N$ | $Q$ | $D_{10-90}$ | $MD$ | $\sigma$ | $MdD$ |
|---|---|---|---|---|---|
| 10 | 5.8 | 1.5 | 4.9 | 3.9 | 5.8 |
| 25 | 6.9 | 1.8 | 5.8 | 4.7 | 6.9 |
| 50 | 7.6 | 2.0 | 6.5 | 5.2 | 7.6 |
| 100 | 8.3 | 2.2 | 7.0 | 5.6 | 8.3 |
| 200 | 8.9 | 2.4 | 7.6 | 6.0 | 8.9 |
| 500 | 9.7 | 2.6 | 8.2 | 6.6 | 9.7 |
| 1,000 | 10.3 | 2.7 | 8.7 | 7.0 | 10.3 |
| 5,000 | 11.5 | 3.0 | 9.8 | 7.8 | 11.5 |
| 10,000 | 12.2 | 3.2 | 10.3 | 8.2 | 12.2 |
| 100,000 | 13.7 | 3.6 | 11.6 | 9.3 | 13.7 |
| 1,000,000 | 15.1 | 4.0 | 12.7 | 10.2 | 15.1 |

The quartile deviation is a somewhat makeshift measure, rather unreliable, not found directly about an average. It is fairly easy to calculate and understand, although more difficult than the range in both these respects. It should be used with the median.

The 10–90 percentile range is high in reliability as compared with other measures obtained from *-ile* points, but lower than the mean and standard deviations. It is a distance extending in both directions from the center, but, as $Q$, may be used with the median. It is also similar to the quartile deviation in relative ease of computation and understanding, but is less frequently employed.

The mean deviation is of above-average reliability. It is somewhat difficult to compute, but readily understood. It is based on all the cases. The mean deviation is unique in that it may be found about either the mean or the median.

The standard deviation is most favored by statisticians, but some of their reasons have slight pertinence to its use in most educational statistical work. It is a function of the normal curve, being the distance from the mean to the point of inflection, that is, the point at which a line tangent to the curve changes direction. It is the most reliable measure of variability, is based upon all the cases, is useful in many other statistical

procedures, and has desirable algebraic relationships and properties. On the other hand, however, it is relatively difficult to compute and the least readily comprehended by persons without statistical training. It gives more weight to the larger deviations than do other variability measures. This may or may not be desirable.

Since the median deviation is regularly found by multiplying the standard deviation by a constant, it shares many of the latter's merits and weaknesses. It is relatively easily understood and remembered because of the 50–50 division point which it represents.

Of these measures, the standard and quartile deviations are far more frequently employed in dealing with educational data than are the others. The mean deviation never has been so used to a great extent and the other two appear to have declined markedly in popularity during recent years.

## The coefficient of variability

All the measures presented in the preceding sections are absolute; therefore they cannot well be employed in comparing the variability of different sets of data except in special cases. The *coefficient of variability* or *of variation* may be used to compare the relative variability of different distributions.

Various formulas for the coefficient of variability have been suggested. All involve the comparison of a measure of variability with a measure of central tendency. This is based on the assumption that the best relative measure of variability is one that compares absolute variability with the size of the measures themselves and that a measure of central tendency is the best measure of their size. The formula most commonly employed is $C$ of $V$, or just $V$, $= 100\frac{\sigma}{M}$. The multiplier 100 is introduced merely to yield a result that is a whole number. For the data often employed previously this formula gives

$$V = 100\,\frac{52.96}{253.64} = 21.$$

The reason why the coefficient of variability is not a thor-

oughly satisfactory measure is that in many instances the scoring system employed does not have a true zero point or equal units. If these two conditions do not hold for a set of data, this or any other similar measure may give misleading information. Despite this weakness, it is the most frequently used measure of relative variability.

### The measurement of skewness

Since the formulas most often employed to measure skewness and kurtosis are based upon measures of central tendency and of variability, it appears appropriate to include them here. The best evidence as to the degree of skewness of a distribution is generally that obtained by inspecting its graphic representation, but a single numerical index is frequently useful. The most commonly employed formula for skewness is

$$Sk = \frac{M - Mo}{\sigma} \text{ or } \frac{3(M - Md)}{\sigma},$$

the latter and generally more convenient form being secured by substituting for $Mo$ its formula, $3Md - 2M$. A positive value indicates that the mean of the data exceeds both their median and their mode, that there is greater bunching of measures in the direction of low scores and greater extension or "tailing out" in that of high ones. A negative value shows just the opposite. The maximum value is $\pm 3.00$, but one greater than $\pm 1.00$ indicates extreme skewness or asymmetry.

Sometimes another formula, generally less satisfactory than the one already given, is employed. It is

$$Sk = \frac{(Q_3 - Md) - (Md - Q_1)}{Q},$$

which by combining terms after removing parentheses may be changed to the more convenient form,

$$Sk = \frac{Q_1 + Q_3 - 2Md}{Q}.$$

Positive and negative values from it have the same general interpretation as those from the other formula. It tends to yield smaller values, which cannot exceed $\pm 2.00$ and do not

often exceed $\pm 1.00$. It appears to be less sensitive to small differences in the shapes of distributions. Except in rare instances, almost always when both are quite near zero, the results from the two formulas applied to the same data have the same sign.

For the 44 cases employed throughout these chapters the first formula gives

$$Sk = \frac{3(253.64 - 242.22)}{52.96} = .65,$$

and the second

$$Sk = \frac{216.00 + 275.00 - 2 \times 242.22}{29.5} = .22.$$

Thus both indicate positive skewness, but the difference between the values they yield is somewhat greater than is usually the case.

## The measurement of kurtosis

The general procedure involved in the simpler measures of kurtosis is to compare a measure of the spread of the central portion of a distribution with that of a larger portion of it. A simple and frequently employed measure of this sort is $Ku = \frac{Q}{D}$. If its value is .2632, the distribution is mesokurtic or normal; if it is larger than that, the distribution is platykurtic; if smaller, it is leptokurtic. It is always positive and ranges from .00 to .50. If it is .00, at least the middle 50 per cent, but not the middle 80 per cent, are bunched at a single point; if it is .3125, the distribution, or at least the middle 80 per cent of it, may form a perfect rectangle; if it is .50, all the cases between the tenth percentile and the first quartile are bunched at one point and all those between the third quartile and the ninetieth percentile at another.

Another less-used measure is $Ku = \frac{Q}{MdD}$. If this is 1.00, the data form a mesokurtic distribution; if it is larger than 1.00, they form a platykurtic one; if less than 1.00, a leptokurtic one. Its lower limit is .00 and its upper limit infinity, but it very rarely exceeds 2.00.

Applied to the data so often used, these formulas give

$$Ku = \frac{29.5}{143.43} = .21$$

and

$$= \frac{29.5}{37.2} = .70,$$

respectively. Since the distribution is far from symmetrical, measures of kurtosis are of doubtful significance for it, but, insofar as they are, the results of both formulas indicate that it is leptokurtic.

## EXERCISES AND PROBLEMS

**1.** Compute the range, the quartile deviation, the 10–90 percentile range, the mean deviation around the mean and the median, the variance, the standard deviation, the median deviation, and the coefficient of variability of the data in each part of Exercises 1, 3, and 6 on pages 69 and 70.

**2.** Compute skewness and kurtosis, by both formulas for each, for the data in each part of Exercises 1 and 3 on page 69.

## REFERENCES


Garrett, Henry E., *Statistics in Psychology and Education.* Second Edition. Pages 59, 113–118. New York: Longmans, Green & Company, Inc., 1937.

Guilford, J. P., *Fundamental Statistics in Psychology and Education.* Chap. IV. New York: McGraw-Hill Book Company, Inc., 1942.

Kenney, John F., *Mathematics of Statistics.* Part I, Chap. V; pages 104–107. New York: D. Van Nostrand Company, Inc., 1939.

Lindquist, E. F., *A First Course in Statistics.* Revised Edition. Chap. VI. Boston: Houghton Mifflin Company, 1942.

Mode, Elmer B., *The Elements of Statistics.* Chap. VIII. New York: Prentice-Hall, Inc., 1941.

Monroe, Walter S., and Engelhart, Max D., *The Scientific Study of Educational Problems.* Pages 74–79, 84. New York: The Macmillan Company, 1926.

Peters, Charles C., and Van Voorhis, Walter R., *Statistical Procedures and Their Mathematical Bases.* Pages 67–82. New York: McGraw-Hill Book Company, Inc., 1940.

Walker, Helen M., *Elementary Statistical Methods.* Chaps. VIII, X; pages 136–139. New York: Henry Holt & Company, Inc., 1943.

CHAPTER VII

# The Coefficient of Correlation

## The meaning of correlation

*Correlation* may be defined as the amount of relationship between paired facts or of the tendency of two or more variables, or attributes, to concomitant variation. To study and find it we must have available two measures of each of a number of individuals. We cannot correlate data from two separate groups of cases.

Since it is often helpful to know the relationship between such characteristics as height and weight, intelligence and achievement, training and success, quality of work in English and in foreign language, and numerous others, correlations are very frequently computed. Although correlation procedures have sometimes been employed when they were not appropriate, they constitute one of the most useful approaches to many problems that arise in the field of education.

There are numerous measures of correlation. Some may be either positive or negative; others are always positive. Positive correlation signifies that as one of the variables or attributes increases, so does the other; negative that as one increases, the other decreases. Most of the measures have values ranging between .00 and 1.00 or $-1.00$ and $+1.00$. Some measure straight-line or rectilinear relationship only; some, curvilinear. Most are for use with variables, but some may be employed with attributes, or with one attribute and one variable. A majority are concerned with two variables or attributes; a few, with more than two.

An important point in connection with all correlation is that in itself it never shows causation. If two series of data are

perfectly correlated, this fact does not prove that either is a cause of the other. It may be that one is, or that some third factor causes both, but proof of such causation must be obtained by other means than correlation. Correlation does suggest the possibility of causal relationships, but no more.

## The nature of the coefficient of correlation

The *coefficient of correlation,* often called the *product-moment coefficient,* is the most commonly used measure of relationship. It is a measure of straight-line relationship of variables, ranging in value from $-1.00$ for perfect negative relationship through .00 for none or pure chance to $+1.00$ for perfect positive. If each measure in one series is connected with the corresponding one in the other by a uniform first-degree algebraic equation, the coefficient of correlation between the two series is 1.00, $+$ or $-$ according to whether the two series increase together or one increases as the other decreases. Another way of stating this is that two series of data correlate perfectly whenever each item in one can be obtained from the corresponding item in the other by adding or subtracting the same amount, multiplying or dividing by the same factor, or any combination thereof, throughout the series.

In order to illustrate more concretely the significance of various values of the coefficient of correlation, Table XIV has been prepared. Each part contains two examples of the value of the coefficient of correlation indicated by its heading. In each example the first column has been headed $X$; the second, $Y$. The two examples in Part A are of perfect positive correlation. In the first each value of the second variable is just 5 less than the corresponding one of the first. In other words, $X = Y + 5$ and $Y = X - 5$, without exception. In the second example of Part A the uniform relationship is not so apparent, but nevertheless exists, since invariably $X = \frac{Y}{2} - 4$ and $Y = 2X + 8$. The examples in Part B show a strong tendency for high values of $X$ to be paired with high values of $Y$, but no uniform relationship exists. In C high values of one variable are as likely to be associated with low as with

TABLE XIV

EXAMPLES OF VARIOUS VALUES OF THE COEFFICIENT OF CORRELATION

| Part A | | | | Part B | | | | Part C | | | | Part D | | | | Part E | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Perfect Positive | | | | High Positive | | | | Low | | | | High Negative | | | | Perfect Negative | | | |
| $X$ | $Y$ | $X$ | $Y$ | $X$ | $Y$ | $X$ | $Y$ | $X$ | $Y$ | $X$ | $Y$ | $X$ | $Y$ | $X$ | $Y$ | $X$ | $Y$ | $X$ | $Y$ |
| 95 | 90 | 26 | 60 | 98 | 99 | 24 | 13 | 75 | 54 | 36 | 54 | 42 | 17 | 57 | 18 | 94 | 29 | 31 | 16 |
| 90 | 85 | 38 | 84 | 97 | 94 | 22 | 17 | 75 | 61 | 85 | 24 | 41 | 19 | 21 | 45 | 86 | 33 | 31 | 16 |
| 85 | 80 | 15 | 38 | 95 | 93 | 35 | 19 | 72 | 42 | 77 | 68 | 37 | 22 | 43 | 28 | 84 | 34 | 33 | 15 |
| 80 | 75 | 21 | 50 | 94 | 92 | 15 | 9 | 70 | 31 | 41 | 73 | 36 | 21 | 52 | 24 | 84 | 34 | 35 | 14 |
| 75 | 70 | 33 | 74 | 91 | 87 | 17 | 7 | 70 | 55 | 28 | 66 | 35 | 24 | 38 | 33 | 80 | 36 | 39 | 12 |
| 70 | 65 | 30 | 68 | 89 | 90 | 21 | 12 | 69 | 27 | 52 | 48 | 31 | 27 | 36 | 35 | 78 | 37 | 41 | 11 |
| 65 | 60 | 28 | 64 | 88 | 91 | 29 | 15 | 64 | 33 | 62 | 53 | 30 | 33 | 27 | 40 | 72 | 40 | 43 | 10 |
| 60 | 55 | 29 | 66 | 88 | 86 | 38 | 21 | 61 | 44 | 53 | 61 | 26 | 38 | 29 | 42 | 70 | 41 | 47 | 8 |
| 55 | 50 | 18 | 44 | 85 | 83 | 17 | 9 | 58 | 34 | 34 | 42 | 24 | 37 | 59 | 16 | 70 | 41 | 47 | 8 |
| 50 | 45 | 24 | 56 | 82 | 79 | 9 | 5 | 55 | 60 | 66 | 70 | 24 | 40 | 46 | 23 | 68 | 42 | 49 | 7 |
| 45 | 40 | 31 | 70 | 78 | 76 | 15 | 7 | 51 | 48 | 21 | 32 | 23 | 42 | 31 | 37 | 64 | 44 | 53 | 5 |
| 40 | 35 | 37 | 82 | 71 | 73 | 23 | 13 | 46 | 50 | 43 | 79 | 18 | 46 | 45 | 26 | 62 | 45 | 55 | 4 |

high values of the other, thus indicating that the coefficients of correlation are low, near .00. Part D presents the reverse of B, a marked tendency for high values of one to be with low values of the other. Finally in E are two examples of perfect negative correlation. In the first $X$ always $= 152 - 2Y$ and $Y = 76 - \frac{X}{2}$, and in the second $X$ always $= 63 - 2Y$ and $Y = 31.5 - \frac{X}{2}$.

Since the coefficient of correlation measures rectilinear relationship, a perfect or 1.00 value thereof means that if the data are plotted they lie on a straight line. If the coefficient is $+1.00$, this line extends from lower left to upper right on a conventional graph, in harmony with the coördinate axis system, and if it is $-1.00$ the line extends from upper left to lower right. If all the data lie on a straight line parallel to either axis, all the values of one variable are the same, and the formula for the coefficient gives an indeterminate value which can be shown to equal .00. As the coefficient becomes progressively less than 1.00, the distribution of the data departs

more from a straight line, until as it approaches .00 the shape assumed approaches a circle, a rectangle, or some irregular form.

The interpretation of coefficients of correlation will be discussed later, but it appears appropriate here to state that from at least two standpoints the coefficient is a minimum measure of relationship. Since it measures rectilinear relationship only, if additional relationship of a curvilinear nature exists, the total relationship is greater than that which is measured. Also, if there are errors in the data, their effect is usually to lower the value obtained, so that it is less than the actual or true value.

Sometimes the fact that the measuring scale used is composed of unequal but systematically changing units produces artificial curvature. In such a case the skew distribution caused by this condition should be transformed into a normal one by the method given in Chapter XIII and then the coefficient of correlation determined.

## Computing coefficients of correlation of ungrouped series

There are many forms in which the formula for the coefficient of correlation may be given. All are reducible to the same expression, hence yield the same result for the same data. Which to employ is a matter of convenience and preference. Only the few which seem most frequently useful will be presented here.

The simplest form is

$$r = \frac{\Sigma xy}{N\sigma_x\sigma_y}$$

or, dividing through by $N$,

$$r = \frac{\frac{\Sigma xy}{N}}{\sigma_x\sigma_y}$$

In this $r$ is the symbol for the coefficient of correlation, $x$ and $y$ are the deviations of the measures from their respective means,[1]

[1] Small letters, the same as the large ones used to denote the measures themselves, may be employed instead of $d$ for deviations from the mean. Thus $x$ and $d_x$ mean the same, $y$ and $d_y$, and so on.

and $\Sigma$, $N$, and $\sigma$ are the same as in previous chapters. If, as usually, assumed instead of true means are used, it becomes

$$r = \frac{\Sigma xy - Nc_xc_y}{N\sigma_x\sigma_y} \text{ or } \frac{\frac{\Sigma xy}{N} - c_xc_y}{\sigma_x\sigma_y}.$$

Since the $c$'s and $\sigma$'s are not actually needed, the formula is sometimes written in terms of the expressions from which they are obtained rather than of them, thus:

$$r = \frac{N\Sigma xy - \Sigma x\Sigma y}{\sqrt{[N\Sigma x^2 - (\Sigma x)^2]\,[N\Sigma y^2 - (\Sigma y)^2]}}$$

or

$$\frac{\Sigma xy - \frac{\Sigma x\Sigma y}{N}}{\sqrt{\left(\Sigma x^2 - \frac{(\Sigma x)^2}{N}\right)\left(\Sigma y^2 - \frac{(\Sigma y)^2}{N}\right)}}.$$

If the assumed mean is taken as 0, a practice sometimes convenient, the formula is in terms of actual scores, $X$ and $Y$, rather than of deviations, and is

$$r = \frac{\Sigma XY - NM_xM_y}{N\sigma_x\sigma_y} \text{ or } \frac{\frac{\Sigma XY}{N} - M_xM_y}{\sigma_x\sigma_y}$$

or

$$\frac{\Sigma XY - NM_xM_y}{\sqrt{(\Sigma X^2 - NM_x^2)(\Sigma Y^2 - NM_y^2)}}$$

if means are employed, and

$$\frac{\Sigma XY - \frac{\Sigma X\Sigma Y}{N}}{\sqrt{\left(\Sigma X^2 - \frac{(\Sigma X)^2}{N}\right)\left(\Sigma Y^2 - \frac{(\Sigma Y)^2}{N}\right)}}$$

if they are not.

The next three tables, Tables XV, XVI, and XVII, present the three most common variations of the formula applied to ungrouped series of data. They differ in where the assumed mean is taken. The entries in the first column of each are $IQ$'s, those in the second per-cent marks on a test. In Table XV the scores themselves are used directly, which is equivalent to assuming the mean at zero. This is frequently known as the Ayres Method. The first two columns contain the scores,

## TABLE XV

COMPUTATION OF COEFFICIENT OF CORRELATION OF UNGROUPED SCORES DIRECTLY FROM SCORES THEMSELVES

| $X$ | $Y$ | $X^2$ | $Y^2$ | $XY$ |
|---|---|---|---|---|
| 135 | 92 | 18225 | 8464 | 12420 |
| 131 | 97 | 17161 | 9409 | 12707 |
| 126 | 95 | 15876 | 9025 | 11970 |
| 119 | 77 | 14161 | 5929 | 9163 |
| 115 | 84 | 13225 | 7056 | 9660 |
| 114 | 89 | 12996 | 7921 | 10146 |
| 108 | 85 | 11664 | 7225 | 9180 |
| 107 | 78 | 11449 | 6084 | 8346 |
| 104 | 83 | 10816 | 6889 | 8632 |
| 104 | 71 | 10816 | 5041 | 7384 |
| 101 | 81 | 10201 | 6561 | 8181 |
| 99 | 76 | 9801 | 5776 | 7524 |
| 98 | 80 | 9604 | 6400 | 7840 |
| 96 | 72 | 9216 | 5184 | 6912 |
| 92 | 75 | 8464 | 5625 | 6900 |
| 92 | 70 | 8464 | 4900 | 6440 |
| 89 | 79 | 7921 | 6241 | 7031 |
| 88 | 73 | 7744 | 5329 | 6424 |
| 86 | 55 | 7396 | 3025 | 4730 |
| 81 | 64 | 6561 | 4096 | 5184 |
| $N = 20 \boxed{2085} = \Sigma X$ | $20 \boxed{1576} = \Sigma Y$ | $20 \boxed{221761} = \Sigma X^2$ | $20 \boxed{126180} = \Sigma Y^2$ | $20 \boxed{166774} = \Sigma XY$ |
| $M_x = 104.25$ | $M_y = 78.80$ | $S_x^2 = 11088.05$ | $S_y^2 = 6309.00$ | 8338.70 |
| | | $M_x^2 = 10868.06$ | $M_y^2 = 6209.44$ | $M_xM_y = 8214.90$ |
| | | $\sigma_x^2 = 219.99$ | $\sigma_y^2 = 99.56$ | $p = 123.80$ |
| | | $\sigma_x = 14.83$ | $\sigma_y = 9.98$ | |

$$r = \frac{123.80}{14.83 \times 9.98} = \frac{123.80}{148.00} = +.84 \text{ or } r = \frac{166774 - 20 \times 104.25 \times 78.80}{\sqrt{(221761 - 20 \times 104.25^2)(126180 - 20 \times 78.80^2)}} = \frac{2476.}{2959.79} = +.84$$

labelled $X$ and $Y$, the next two, headed $X^2$ and $Y^2$, their squares, and the last or $XY$ column their products. All are totalled, and the sums divided by $N$, here 20. The results for the first two columns are the means, 104.25 for the $X$ scores and 78.80 for the $Y$ scores. The steps at the bottoms of the next two columns are the same as on page 89, with the apparent but not real difference that $M^2$ instead of $c^2$ is subtracted to give $\sigma^2$. It is not real because in this case, with the assumed mean at 0, $M = c$. Therefore the two standard deviations, 14.83 and 9.98, are found as on the page just mentioned. Finally from the sum of the entries in the last, or XY, column divided by $N$ the product of the two means is subtracted. Here $\Sigma XY = 166{,}774$, which divided by 20 gives 8,338.70. This less $M_xM_y$, 8,214.90, leaves 123.80, labelled $p$, which is the symbol sometimes used for the numerator of the formula.

At the foot of the table $r$ is found by two variations of the formula. First, using

$$r = \frac{\frac{\Sigma XY}{N} - M_xM_y}{\sigma_x\sigma_y},\ r = \frac{123.80}{14.83 \times 9.98} = +.84.$$

Second, employing the next to the last form,

$$r = \frac{166774 - 20 \times 104.25 \times 78.80}{\sqrt{(221761 - 20 \times 104.25^2)(126180 - 20 \times 78.80^2)}} = +.84.$$

Any of the others may also be used, with the same result.

It is worth noting that the numerator of the final fraction obtained from the second variation of the formula is $N$ times that obtained from the first variation employed. Thus here, $2{,}476 = 20 \times 123.80$. Likewise the denominators exhibit the same relationship. Here, 2,959.79 is not exactly $20 \times 148.00$ or 2,960 because some decimals have been dropped. Furthermore, each $\sigma$ equals the square root of the corresponding term in parenthesis in the denominator divided by $N$. In this case,

$$14.83 = \sqrt{\frac{221761 - 20 \times 104.25^2}{20}}$$

and

$$9.98 = \sqrt{\frac{126180 - 20 \times 78.80^2}{20}}.$$

## TABLE XVI

COMPUTATION OF COEFFICIENT OF CORRELATION OF UNGROUPED SCORES WITH ASSUMED MEANS NEAR LOWEST SCORES

| $X$ | $Y$ | $x$ | $y$ | $x^2$ | $y^2$ | $xy$ |
|---|---|---|---|---|---|---|
| 135 | 92 | 55 | 42 | 3025 | 1764 | 2310 |
| 131 | 97 | 51 | 47 | 2601 | 2209 | 2397 |
| 126 | 95 | 46 | 45 | 2116 | 2025 | 2070 |
| 119 | 77 | 39 | 27 | 1521 | 729 | 1053 |
| 115 | 84 | 35 | 34 | 1225 | 1156 | 1190 |
| 114 | 89 | 34 | 39 | 1156 | 1521 | 1326 |
| 108 | 85 | 28 | 35 | 784 | 1225 | 980 |
| 107 | 78 | 27 | 28 | 729 | 784 | 756 |
| 104 | 83 | 24 | 33 | 576 | 1089 | 792 |
| 104 | 71 | 24 | 21 | 576 | 441 | 504 |
| 101 | 81 | 21 | 31 | 441 | 961 | 651 |
| 99 | 76 | 19 | 26 | 361 | 676 | 494 |
| 98 | 80 | 18 | 30 | 324 | 900 | 540 |
| 96 | 72 | 16 | 22 | 256 | 484 | 352 |
| 92 | 75 | 12 | 25 | 144 | 625 | 300 |
| 92 | 70 | 12 | 20 | 144 | 400 | 240 |
| 89 | 79 | 9 | 29 | 81 | 841 | 261 |
| 88 | 73 | 8 | 23 | 64 | 529 | 184 |
| 86 | 55 | 6 | 5 | 36 | 25 | 30 |
| 81 | 64 | 1 | 14 | 1 | 196 | 14 |
| $AssM_x = 80.$ | $AssM_y = 50.$ | 20 $\boxed{485} = \Sigma x$ | 20 $\boxed{576} = \Sigma y$ | 20 $\boxed{16161} = \Sigma x^2$ | 20 $\boxed{18580} = \Sigma y^2$ | 20 $\boxed{16444} = \Sigma xy$ |
| | | $c_x = 24.25$ | $c_y = 28.80$ | $S_x^2 = 808.05$ | $S_y^2 = 929.00$ | 822.20 |
| | | | | $c_x^2 = 588.06$ | $c_y^2 = 829.44$ | $c_x c_y = 698.40$ |
| | | | | $\sigma_x^2 = 219.99$ | $\sigma_y^2 = 99.56$ | $p = 123.80$ |
| | | | | $\sigma_x = 14.83$ | $\sigma_y = 9.98$ | |

$$r = \frac{123.80}{14.83 \times 9.98} = \frac{123.80}{148.00} = +.84 \text{ or } r = \frac{16444 - \dfrac{485 \times 576}{20}}{\sqrt{\left(16161 - \dfrac{485^2}{20}\right)\left(18580 - \dfrac{576^2}{20}\right)}} = \frac{2476.}{2959.79} = +.84$$

The same relationships also hold for the data in Tables XVI and XVII.

Table XVI presents the same computation with the use of assumed means near the smallest scores. In this method each assumed mean is taken as the smallest score in the series or some "round" number slightly smaller. In this case, since the smallest $X$ value is 81, the assumed mean of the $X$'s is taken as 80 and, since the smallest $Y$ is 55, the assumed mean of the $Y$'s as 50. The column headed $x$ contains the remainders after subtracting 80 from each $X$ score, and that headed $y$ the remainders after subtracting 50 from each $Y$ score. The next two contain the squares of the $x$'s and $y$'s, and the last contain their products. The same steps are carried out as for the data in Table XV. giving the same result,

$$r = \frac{123.80}{14.83 \times 9.98} = +.84.$$

Also, the last of the formulas above for use with deviations is employed, with the totals of the five columns. It gives

$$r = \frac{16444 - \dfrac{485 \times 576}{20}}{\sqrt{\left(16161 - \dfrac{485^2}{20}\right)\left(18580 - \dfrac{576^2}{20}\right)}} = +.84.$$

The advantage of the method just described is that the sizes of the numbers dealt with are much reduced. Whether this reduction is sufficient to pay for the additional step of subtracting the assumed means from the actual scores depends upon their relative sizes and the relative facility of the computer in handling numbers of different sizes. If the range of scores is much greater than the smallest one, so that the ratio of reduction by subtracting assumed means is small, so doing is not worth while. For example, if the $X$ scores ranged from 81 to 535 instead of to 135, subtracting 80 would hardly be worth the trouble. Similarly, if they ranged from 21, instead of 81, to 135, subtracting 20 would not pay.

Table XVII illustrates the method in which the assumed means are taken close to the true means. This has the advantage of giving even smaller numbers than the previous method,

TABLE XVII

COMPUTATION OF COEFFICIENT OF CORRELATION OF UNGROUPED SCORES WITH ASSUMED MEANS NEAR TRUE MEANS

| $X$ | $Y$ | $x$ | $y$ | $x^2$ | $y^2$ | $xy$ |
|---|---|---|---|---|---|---|
| 135 | 92 | +35 | +12 | 1225 | 144 | +420 |
| 131 | 97 | +31 | +17 | 961 | 289 | +527 |
| 126 | 95 | +26 | +15 | 676 | 225 | +390 |
| 119 | 77 | +19 | −3 | 361 | 9 | −57 |
| 115 | 84 | +15 | +4 | 225 | 16 | +60 |
| 114 | 89 | +14 | +9 | 196 | 81 | +126 |
| 108 | 85 | +8 | +5 | 64 | 25 | +40 |
| 107 | 78 | +7 | −2 | 49 | 4 | −14 |
| 104 | 83 | +4 | +3 | 16 | 9 | +12 |
| 104 | 71 | +4 | −9 | 16 | 81 | −36 |
| 101 | 81 | +1 | +1 | 1 | 1 | +1 |
| 99 | 76 | −1 | −4 | 1 | 16 | +4 |
| 98 | 80 | −2 | 0 | 4 | 0 | 0 |
| 96 | 72 | −4 | −8 | 16 | 64 | +32 |
| 92 | 75 | −8 | −5 | 64 | 25 | +40 |
| 92 | 70 | −8 | −10 | 64 | 100 | +80 |
| 89 | 79 | −11 | −1 | 121 | 1 | +11 |
| 88 | 73 | −12 | −7 | 144 | 49 | +84 |
| 86 | 55 | −14 | −25 | 196 | 625 | +350 |
| 81 | 64 | −19 | −16 | 361 | 256 | +304 |
| $AssM_x = 100.$ | $AssM_y = 80.$ | +164 | +66 | $20\ \boxed{4761} = \Sigma x^2$ | $20\ \boxed{2020} = \Sigma y^2$ | +2481 |
| | | −79 | −90 | $S_x^2 = 238.05$ | $S_y^2 = 101.00$ | −107 |
| | | $20\ \boxed{+85} = \Sigma x$ | $20\ \boxed{-24} = \Sigma y$ | $c_x^2 = 18.06$ | $c_y^2 = 1.44$ | $20\ \boxed{+2374} = \Sigma xy$ |
| | | $c_x = +4.25$ | $c_y = -1.2$ | $\sigma_x^2 = 219.99$ | $\sigma_y^2 = 99.56$ | 118.70 |
| | | | | $\sigma_x = 14.83$ | $\sigma_y = 9.98$ | $c_x c_y = -5.10$ |
| | | | | | | $p = 123.80$ |

$$r = \frac{123.80}{14.83 \times 9.98} = \frac{123.80}{148.00} = +.84 \text{ or } r = \frac{2374 - \frac{85 \times (-24)}{20}}{\sqrt{\left(4761 - \frac{85^2}{20}\right)\left(2020 - \frac{24^2}{20}\right)}} = \frac{2476.}{2959.79} = +.84$$

but the disadvantage of introducing the minus signs. It is, except for the last column and the final steps, the same as the short method of computing the mean, given on page 56, and of finding the standard deviation, on page 89. In form it is just the same as that in Table XVI. The same two variations of the formula are employed, with the same result.

Sometimes all the scores in a series can be evenly divided by the same factor. If so, such division may be worth while. It is merely a question of whether the reduction in size is sufficient to outweigh the extra labor of the additional step. If all end in one or more 0's which may be dropped, it is evidently worth doing so, but otherwise it may or may not be.

## Computing coefficients of correlation of grouped series

The process of computing coefficients of correlation of grouped series is the same in principle as that for ungrouped series, but must be carried out in different form. The same principles that determine whether or not a single series should be dealt with in ungrouped or grouped form apply in cases of correlation of two series. Generally, if there are more than 30 or 40 cases, they are grouped; if there are less, they are not. The resulting table is known as a *correlation table*, a *table of double entry*, or a *double frequency table*. It is constructed in harmony with the coördinate axis system. The scale for one series, usually the first, is laid off upon the $X$ or horizontal axis; that for the other, upon the $Y$ or vertical axis. A table is thus formed, composed of a number of cells or compartments, each of which contains a number showing how many cases have the $Y$ value of the row and $X$ value of the column which intersect to form the cell. The $X$ values are commonly indicated by figures or other symbols, called *column headings* or *captions*, at the top of the table; the $Y$ values, by others, called *stubs*, at the left.

To prepare such a table, the worker first should examine each series and, in accord with the suggestions given in Chapter II, decide how to classify it. One additional consideration is that the numbers of classes in the two series correlated should be approximately the same. Having done this, he should prepare

the table, insert the captions and stubs, and tabulate the data. Probably the most convenient method of doing the latter is to place a tally mark, made with a fairly soft pencil, in the proper square for each pair of numbers. After doing this once, he should repeat the process and compare results, to eliminate probability of error. If results agree—or after they are corrected, if they do not—an entry showing the number of tally marks in each cell should be entered therein. If entry is made with ink or a hard pencil, the tally marks can easily be erased without disturbing it, and a clean, neat table left without the labor of preparing one and then recopying it.

To illustrate the construction of a correlation table, and later the computation of the coefficient thereof, the data in Table XVIII may be employed. The first series are the total scores

TABLE XVIII

Scores of a Class on a Series of Short Tests and on the Final Examination*

| Short Tests | Final | Short Tests | Final | Short Tests | Final | Short Tests | Final | Short Tests | Final |
|---|---|---|---|---|---|---|---|---|---|
| 336 | 130 | 291 | 116 | 259 | 115 | 200 | 95 | 128 | 76 |
| 321 | 114 | 272 | 108 | 186 | 96 | 198 | 96 | 353 | 126 |
| 226 | 102 | 258 | 111 | 222 | 95 | 209 | 102 | 267 | 112 |
| 270 | 109 | 202 | 93 | 174 | 79 | 143 | 78 | 326 | 119 |
| 333 | 131 | 250 | 109 | 315 | 119 | 126 | 75 | 267 | 104 |
| 351 | 132 | 251 | 105 | 230 | 98 | 291 | 118 | 182 | 94 |
| 274 | 106 | 339 | 125 | 215 | 100 | 325 | 128 | 252 | 100 |
| 175 | 94 | 269 | 100 | 317 | 129 | 235 | 106 | 175 | 95 |
| 202 | 99 | 165 | 94 | 296 | 124 | 238 | 101 | 263 | 118 |
| 241 | 101 | 237 | 116 | 292 | 109 | 183 | 106 | 184 | 107 |

* These scores form the basis of Tables XIX, XX, and XXI.

of 50 students on a series of short tests; the second, their scores on the final examination. For example, one student had a total of 336 points on the short tests and made 130 on the final, another had 321 and made 114, and so on.

Table XIX shows the initial form of a correlation table for the data in Table XVIII. Classes with an interval of 20, beginning at 120, have been chosen for the total scores and with an interval of 5, beginning at 75, for the final examination

scores. A tally mark for each pair of scores is then placed in the proper cell. Thus for the first pair, 336 and 130, such a mark has been placed in the cell formed by the intersection of the 320-column and the 130-row; for the next, 321 and 114, in

TABLE XIX

First Step in Construction of Correlation Table Containing Data in Table XVIII

| Final Exam. Score | Total Score on Short Tests | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 120– | 140– | 160– | 180– | 200– | 220– | 240– | 260– | 280– | 300– | 320– | 340– |
| 130– | | | | | | | | | | | // | / |
| 125– | | | | | | | | | | / | // | / |
| 120– | | | | | | | | | / | | | |
| 115– | | | | | | / | / | / | // | / | / | |
| 110– | | | | | | | / | / | | | / | |
| 105– | | | | // | | / | // | /// | / | | | |
| 100– | | | | | // | // | // | // | | | | |
| 95– | | | / | // | // | // | | | | | | |
| 90– | | | // | / | / | | | | | | | |
| 85– | | | | | | | | | | | | |
| 80– | | | | | | | | | | | | |
| 75– | // | / | / | | | | | | | | | |

that made by the intersection of the 320-column and the 110-row; and so on until all are entered. This has been checked and the final result is as shown in the table.

These data have been entered in the next two tables, but with figures instead of tally marks in the cells. Also, they have been totalled both ways and the results entered in the total or frequency column and row. When this has been done, they are ready for the computation of the coefficient of correlation between the two series. In Table XX this has been done by

the use of assumed means just below the lowest scores, that is, by the so-called Ayres Method. As applied to a correlation table, it provides that the deviations assigned the classes begin with 1 for the lowest class and increase by one per class until the largest equals the number of classes. In this instance, since each variable is grouped in twelve classes, both $x$ and $y$ range from 1 up to 12.

The next two rows, labelled $fx$ and $fx^2$, and the next two columns, headed $fy$ and $fy^2$, are the same as in the computation of the standard deviation shown on page 90, except that there an assumed mean near the true mean is used. In other words, each $fx$ entry is the product of the $f$ and $x$ entries in the same column or row, and each $fx^2$ entry the product of the corresponding $x$ and $fx$ entries, and similarly for $y$.

The $\Sigma x$ column is new. Each entry therein is the sum of the products of the entries in that row of the table times the $x$ values of their respective columns. Thus, for the first row,

$$2 \times 11 + 1 \times 12 = 34;$$

for the second,

$$1 \times 10 + 2 \times 11 + 1 \times 12 = 44;$$

and similarly for the others. The total of the $\Sigma x$ column, here 341, should equal the total of the $fx$ row. For the $\Sigma xy$ column each entry in the $\Sigma x$ column is multiplied by the $y$ value of the same row. Thus $12 \times 34 = 408$, the first such product; $11 \times 44 = 484$, the second; and so forth. This column is then summed, the result, in this instance, being 2,701.

The two rows below the wavy line are not necessary, but serve to check part of the work. The first, labelled $\Sigma y$, contains entries computed similarly to those in the $\Sigma x$ column, but vertically rather than horizontally. For example, the $\Sigma y$ entry for the 180-column is 28, the sum of

$$2 \times 7 + 2 \times 5 + 1 \times 4.$$

The total of the $\Sigma y$ row should equal that of the $fy$ column. Here both are 343. Each $\Sigma y$ entry is multiplied by the corresponding $x$, the result entered in the $\Sigma xy$ row, and the total found. It should, of course, equal that of the $\Sigma xy$ column, in this case 2,701.

TABLE

COMPUTATION OF COEFFICIENT OF CORRELATION OF DATA IN

| Final Exam. Score | Total Score on | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | 120– | 140– | 160– | 180– | 200– | 220– | 240– | 260– | 280– |
| 130– | | | | | | | | | |
| 125– | | | | | | | | | |
| 120– | | | | | | | | | 1 |
| 115– | | | | | | 1 | 1 | 1 | 2 |
| 110– | | | | | | | 1 | 1 | |
| 105– | | | | 2 | | 1 | 2 | 3 | 1 |
| 100– | | | | | 2 | 2 | 2 | 2 | |
| 95– | | | 1 | 2 | 2 | 2 | | | |
| 90– | | | 2 | 1 | 1 | | | | |
| 85– | | | | | | | | | |
| 80– | | | | | | | | | |
| 75– | 2 | 1 | 1 | | | | | | |
| $f_x$ | 2 | 1 | 4 | 5 | 5 | 6 | 6 | 7 | 4 |
| $x$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| $fx$ | 2 | 2 | 12 | 20 | 25 | 36 | 42 | 56 | 36 |
| $fx^2$ | 2 | 4 | 36 | 80 | 125 | 216 | 294 | 448 | 324 |
| $\Sigma y$ | 2 | 1 | 14 | 28 | 26 | 38 | 43 | 50 | 35 |
| $\Sigma xy$ | 2 | 2 | 42 | 112 | 130 | 228 | 301 | 400 | 315 |

$$r = \frac{2701 - \dfrac{341 \times 343}{50}}{\sqrt{\left(2743 - \dfrac{341^2}{50}\right)\left(2747 - \dfrac{343^2}{50}\right)}} = \frac{361.74}{405.53} = +.89$$

XX

CORRELATION TABLE WITH ASSUMED MEANS BELOW LOWEST SCORES

| Short Tests | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 300– | 320– | 340– | $f_y$ | $y$ | $fy$ | $fy^2$ | $\Sigma x$ | $\Sigma xy$ |
| | 2 | 1 | 3 | 12 | 36 | 432 | 34 | 408 |
| 1 | 2 | 1 | 4 | 11 | 44 | 484 | 44 | 484 |
| | | | 1 | 10 | 10 | 100 | 9 | 90 |
| 1 | 1 | | 7 | 9 | 63 | 567 | 60 | 540 |
| | 1 | | 3 | 8 | 24 | 192 | 26 | 208 |
| | | | 9 | 7 | 63 | 441 | 61 | 427 |
| | | | 8 | 6 | 48 | 288 | 52 | 312 |
| | | | 7 | 5 | 35 | 175 | 33 | 165 |
| | | | 4 | 4 | 16 | 64 | 15 | 60 |
| | | | 0 | 3 | 0 | 0 | 0 | 0 |
| | | | 0 | 2 | 0 | 0 | 0 | 0 |
| | | | 4 | 1 | 4 | 4 | 7 | 7 |
| 2 | 6 | 2 | 50 | | 343 | 2747 | 341 | 2701 |
| 10 | 11 | 12 | | | | | | |
| 20 | 66 | 24 | 341 | | | | | |
| 200 | 726 | 288 | 2743 | | | | | |
| 20 | 63 | 23 | 343 | | | | | |
| 200 | 693 | 276 | 2701 | | | | | |

$$c_x = \frac{341}{50} = 6.82 \qquad c_y = \frac{343}{50} = 6.86$$

$$S_x^2 = \frac{2743}{50} = 54.8600 \qquad S_y^2 = \frac{2747}{50} = 54.9400$$

$$c_x^2 = 46.5124 \qquad c_y^2 = 47.0596$$

$$\sigma_x^2 = 8.3476 \qquad \sigma_y^2 = 7.8804$$

$$\sigma_x = 2.889 \qquad \sigma_y = 2.807$$

$$\frac{2701}{50} = 54.0200$$

$$c_x c_y = 46.7852$$

$$p = 7.2348$$

$$r = \frac{7.2348}{2.889 \times 2.807} = \frac{7.2348}{8.1094} = +.89$$

After the totals of the $fx$ and $fx^2$ rows, of the $fy$ and $fy^2$ columns, and of $\Sigma xy$ have been obtained, the coefficient of correlation may, as in the case of ungrouped series, be computed by any of a number of variations of the general formula. They are the same as those previously given for ungrouped series, but the work is done in terms of deviations, expressed in intervals, rather than in either actual scores or deviations expressed in original units. In Table XX the computation of $r$ by two methods, one involving the finding of the corrections and standard deviations, and the other not, is shown.

The method used at the lower right of the table is based upon the fourth formula for $r$ given above. The corrections and standard deviations are found, as back in Table X, giving $c_x = 6.82$, $c_y = 6.86$, $\sigma_x = 2.889$, and $\sigma_y = 2.807$, all in terms of intervals. Since $\Sigma xy = 2701$,

$$\frac{\Sigma xy}{N} = \frac{2701}{50} = 54.0200.$$

Subtracting $c_x c_y$ from this leaves 7.2348, the numerator of the final fraction.

$$r = \frac{7.2348}{2.889 \times 2.807} = .89.$$

Two points not essential to this computation appear worth mentioning in connection with it. If the means are desired, they may be found by the usual formula, $M = AssM + ci$. In this case the assumed mean is always the mid-point of the next lower class below the lowest one in the table. Therefore,

$$M_x = 110. + 6.82 \times 20. = 246.4$$

and

$$M_y = 72.5 + 6.86 \times 5 = 106.8.$$

Secondly, if the standard deviations are to be reported as such, they should be in terms of original or score units, not of intervals as they are employed in the calculation. That is,

$$\sigma_x = 2.889 \times 20. = 57.78$$

and

$$\sigma_y = 2.807 \times 5. = 14.035.$$

The other of the two methods of finding the coefficient of correlation illustrated in Table XX employs the second of the formulas used when the $c$'s and $\sigma$'s are not needed. Substituting the proper totals in this formula gives

$$r = \frac{2701 - \dfrac{341 \times 343}{50}}{\sqrt{\left(2743 - \dfrac{341^2}{50}\right)\left(2747 - \dfrac{343^2}{50}\right)}} = +.89,$$

as by the other method.

It is true here, just as for Tables XV, XVI, and XVII, that the numerator and denominator of the fraction yielded by the second method are each $N$, in this instance 50, times those of the fraction resulting from the first method. Also, in both places, each $\sigma$ is the square root of the corresponding quantity in parenthesis divided by $N$.

Table XXI contains the same data and the computation of their coefficient of correlation by the same two formulas used in Table XX, but with assumed means taken near the true means. Since some of the steps are the same as those for means and standard deviations, and all are the same as those in Table XX, they are not explained here. Both formulas yield the same final fraction and value of $r$ as in the previous table.

A number of so-called correlation charts, really work sheets, intended to assist those who are computing coefficients from correlation tables, have been published. They regularize procedure by providing a specific place for each step, often with explanation right at hand, but the writer does not believe that they offer much advantage over the use of cross-section paper accompanied by the habit of placing each step in the same position.

Grouping data to be correlated sacrifices accuracy for convenience, just as any grouping does. It tends to yield lower coefficients than would be obtained from the exact data. If the series correlated are not very dissimilar in shape and if the number of classes in each is ten or greater, the inaccuracy is generally small. A study by the writer indicated that with

TABLE

COMPUTATION OF COEFFICIENT OF CORRELATION OF DATA IN

| Final Exam. Score | Total Score on | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | 120– | 140– | 160– | 180– | 200– | 220– | 240– | 260– | 280– |
| 130– | | | | | | | | | |
| 125– | | | | | | | | | |
| 120– | | | | | | | | | 1 |
| 115– | | | | | | 1 | 1 | 1 | 2 |
| 110– | | | | | | | 1 | 1 | |
| 105– | | | | 2 | | 1 | 2 | 3 | 1 |
| 100– | | | | | 2 | 2 | 2 | 2 | |
| 95– | | | 1 | 2 | 2 | 2 | | | |
| 90– | | | 2 | 1 | 1 | | | | |
| 85– | | | | | | | | | |
| 80– | | | | | | | | | |
| 75– | 2 | 1 | 1 | | | | | | |
| $f_x$ | 2 | 1 | 4 | 5 | 5 | 6 | 6 | 7 | 4 |
| $x$ | −6 | −5 | −4 | −3 | −2 | −1 | 0 | +1 | +2 |
| $fx$ | −12 | −5 | −16 | −15 | −10 | −6 | −64 | +7 | +8 |
| $fx^2$ | 72 | 25 | 64 | 45 | 20 | 6 | 0 | 7 | 16 |
| $\Sigma y$ | −12 | −6 | −14 | −7 | −9 | −4 | +1 | +1 | +7 |
| $\Sigma xy$ | +72 | +30 | +56 | +21 | +18 | +4 | 0 | +1 | +14 |

$$r = \frac{363 - \dfrac{(-9)(-7)}{50}}{\sqrt{\left(419 - \dfrac{9^2}{50}\right)\left(395 - \dfrac{7^2}{50}\right)}} = \frac{361.74}{405.53} = +.89$$

XXI

CORRELATION TABLE WITH ASSUMED MEANS NEAR TRUE MEANS

| Short Tests | | | | $y$ | $fy$ | $fy^2$ | $\Sigma x$ | $\Sigma xy$ |
|---|---|---|---|---|---|---|---|---|
| 300– | 320– | 340– | $f_y$ | | | | | |
| | 2 | 1 | 3 | +5 | +15 | 75 | +13 | +65 |
| 1 | 2 | 1 | 4 | +4 | +16 | 64 | +16 | +64 |
| | | | 1 | +3 | +3 | 9 | +2 | +6 |
| 1 | 1 | | 7 | +2 | +14 | 28 | +11 | +22 |
| | 1 | | 3 | +1 | +3 | 3 | +5 | +5 |
| | | | 9 | 0 | +51 | 0 | −2 | 0 |
| | | | 8 | −1 | −8 | 8 | −4 | +4 |
| | | | 7 | −2 | −14 | 28 | −16 | +32 |
| | | | 4 | −3 | −12 | 36 | −13 | +39 |
| | | | 0 | −4 | 0 | 0 | 0 | 0 |
| | | | 0 | −5 | 0 | 0 | 0 | 0 |
| | | | 4 | −6 | −24 | 144 | −21 | +126 |
| 2 | 6 | 2 | 50 | | −58 | 395 | +47 | +363 |
| +3 | +4 | +5 | | | −7 | | −56<br>− 9 | |
| +6 | +24 | +10 | +55 \| − 9 | | | | | |
| 18 | 96 | 50 | 419 | | | | | |
| +6 | +21 | +9 | −52\|+45\|−7 | | | | | |
| +18 | +84 | +45 | +363 | | | | | |

$$c_x = \frac{-9}{50} = -.18 \qquad c_y = \frac{-7}{50} = -.14$$

$$S_x^2 = \frac{419}{50} = 8.3800 \qquad S_y^2 = \frac{395}{50} = 7.9000$$

$$c_x^2 = .0324 \qquad c_y^2 = .0196$$

$$\sigma_x^2 = 8.3476 \qquad \sigma_y^2 = 7.8804$$

$$\sigma_x = 2.889 \qquad \sigma_y = 2.807$$

$$\frac{363}{50} = 7.2600$$

$$c_x c_y = .0252$$

$$p = 7.2348$$

$$r = \frac{7.2348}{2.889 \times 2.807} = \frac{7.2348}{8.1094} = +.89$$

only five classes in each variable over half the differences are less than .05; that with ten classes in each over two-thirds are less than .02; and that with twenty classes about three-fourths are less than .01.

A correction for grouping may be obtained by substituting the corrected value of each $\sigma$, as given on page 91, for the actually obtained value. The formula given there, for $\sigma$ expressed in intervals, was

$$\sigma_{corr} = \sqrt{\sigma^2_{obt} - \frac{1}{12}}.$$

Employing this for the data in Table XXI gives

$$\sigma_{x_{corr}} = \sqrt{2.889^2 - .0833} = 2.875$$

and

$$\sigma_{y_{corr}} = \sqrt{2.807^2 - .0833} = 2.792,$$

whence

$$r_{corr} = \frac{7.2348}{2.875 \times 2.792} = +.90,$$

an increase of only .01 from +.89, the uncorrected value.

## Computing intercorrelations among a number of variables

The various intercorrelations among a number of variables are often needed. These must be based upon the same cases in each; hence their intercorrelations involve repeated use of the same totals, means, corrections, standard deviations, and so forth. Only the products of the pairs of variables are unique to each correlation. Therefore anyone who is computing such intercorrelations should so plan his work as to avoid unnecessary duplication. Table XXII illustrates how this may be done for ungrouped series. The first four columns, headed $A$, $E$, $G$, and $I$, contain the school marks of 20 pupils in algebra, English, general science, and industrial arts. They were given in terms of a five-letter marking system, but have been transmuted into numbers from 5 to 1. The next four columns contain their squares; and the last six, the various products. These are then totalled and the means, standard deviations,

TABLE XXII

COMPUTATION OF INTERCORRELATIONS AMONG A NUMBER OF SERIES OF UNGROUPED VARIABLES

| $A$ | $E$ | $G$ | $I$ | $A^2$ | $E^2$ | $G^2$ | $I^2$ | $AE$ | $AG$ | $AI$ | $EG$ | $EI$ | $GI$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 4 | 4 | 5 | 5 | 16 | 16 | 25 | 25 | 16 | 20 | 20 | 20 | 20 | 25 |
| 2 | 2 | 3 | 2 | 4 | 4 | 9 | 4 | 4 | 6 | 4 | 6 | 4 | 6 |
| 3 | 4 | 4 | 4 | 9 | 16 | 16 | 16 | 12 | 12 | 12 | 16 | 16 | 16 |
| 1 | 2 | 2 | 3 | 1 | 4 | 4 | 9 | 2 | 2 | 3 | 4 | 6 | 6 |
| 2 | 3 | 1 | 2 | 4 | 9 | 1 | 4 | 6 | 2 | 4 | 3 | 6 | 2 |
| 3 | 3 | 3 | 3 | 9 | 9 | 9 | 9 | 9 | 9 | 9 | 9 | 9 | 9 |
| 3 | 3 | 3 | 4 | 9 | 9 | 9 | 16 | 9 | 9 | 12 | 9 | 12 | 12 |
| 4 | 5 | 4 | 4 | 16 | 25 | 16 | 16 | 20 | 16 | 16 | 20 | 20 | 16 |
| 2 | 2 | 3 | 3 | 4 | 4 | 9 | 9 | 4 | 6 | 6 | 6 | 6 | 9 |
| 2 | 4 | 3 | 3 | 4 | 16 | 9 | 9 | 8 | 6 | 6 | 12 | 12 | 9 |
| 5 | 3 | 5 | 5 | 25 | 9 | 25 | 25 | 15 | 25 | 25 | 15 | 15 | 25 |
| 3 | 3 | 2 | 4 | 9 | 9 | 4 | 16 | 9 | 6 | 12 | 6 | 12 | 8 |
| 3 | 4 | 3 | 3 | 9 | 16 | 9 | 9 | 12 | 9 | 9 | 12 | 12 | 9 |
| 1 | 2 | 1 | 1 | 1 | 4 | 1 | 1 | 2 | 1 | 1 | 2 | 2 | 1 |
| 3 | 3 | 2 | 3 | 9 | 9 | 4 | 9 | 9 | 6 | 9 | 6 | 9 | 6 |
| 3 | 4 | 3 | 3 | 9 | 16 | 9 | 9 | 12 | 9 | 9 | 12 | 12 | 9 |
| 2 | 1 | 2 | 4 | 4 | 1 | 4 | 16 | 2 | 4 | 8 | 2 | 4 | 8 |
| 4 | 3 | 4 | 5 | 16 | 9 | 16 | 25 | 12 | 16 | 20 | 12 | 15 | 20 |
| 3 | 2 | 3 | 3 | 9 | 4 | 9 | 9 | 6 | 9 | 9 | 6 | 6 | 9 |
| 1 | 3 | 2 | 2 | 1 | 9 | 4 | 4 | 3 | 2 | 2 | 6 | 6 | 4 |
| 20 54 | 60 | 58 | 66 | 168 | 198 | 192 | 240 | 172 | 175 | 196 | 184 | 204 | 209 |
| $M = 2.7$ | 3.0 | 2.9 | 3.3 | $S^2 = 8.40$ | 9.90 | 9.60 | 12.00 | 8.60 | 8.75 | 9.80 | 9.20 | 10.20 | 10.45 |
| | | | | $M^2 = 7.29$ | 9.00 | 8.41 | 10.89 | $MM = 8.10$ | 7.83 | 8.91 | 8.70 | 9.90 | 9.57 |
| | | | | $\sigma^2 = 1.11$ | .90 | 1.19 | 1.11 | $p = .50$ | .92 | .89 | .50 | .30 | .88 |
| | | | | $\sigma = 1.05$ | .95 | 1.09 | 1.05 | | | | | | |

$$r_{AE} = \frac{.50}{1.05 \times .95} = +.50 \qquad r_{AG} = \frac{.92}{1.05 \times 1.09} = +.80 \qquad r_{AI} = \frac{.89}{1.05 \times 1.05} = +.81$$

$$r_{EG} = \frac{.50}{.95 \times 1.09} = +.48 \qquad r_{EI} = \frac{.30}{.95 \times 1.05} = +.30 \qquad r_{GI} = \frac{.88}{1.09 \times 1.05} = +.77$$

numerators of the final fractions, and finally the values of the six possible correlation coefficients found by the method first employed in Table XV. The resulting coefficients, given at the bottom of Table XXII, show a range in values of from +.30 up to +.81.

## Limitations upon the use of coefficients of correlation

It has already been stated that coefficients of correlation measure only rectilinear relationship, that variable errors in the data tend to cause the values computed to be less than those derived from perfectly accurate scores, and that they never prove causative relationship. In addition, there are certain other limitations and cautions that should be heeded.

One of these has to do with averaging. The ordinary mean of two or more correlation coefficients is not at all sure to be very close to the value that would be obtained if the data from which the two or more are derived were all combined and a single coefficient determined. If the means and standard deviations of each variable, as it appears in the several correlations, and also the numbers of cases, are each approximately the same, then the mean is probably close to the value for all the data combined. Two methods have been suggested as better, but since one regularly yields results larger than the mean, and the other smaller, it is doubtful which, if either, should be employed. Probably the better of the two calls for taking the square root of the mean of the squares of the coefficients to be averaged. Thus for $r$'s of +.48, +.52, +.61, and +.75, the result is

$$\sqrt{\frac{.48^2 + .52^2 + .61^2 + .75^2}{4}} = +.60.$$

If some of the $r$'s are positive and some negative, this method is not valid.

What is called spurious correlation arises through the effect of an irrelevant factor upon the variables being correlated. In such cases, the computed coefficient is correct mathematically, but liable to misinterpretation because of neglect of the irrelevant factor. For example, if total scores on a test are

correlated with those on one of its subtests, the fact that the subtest is contained in the total serves to increase the obtained coefficient. In general, if a subtest is uncorrelated with the other subtests composing the whole test, the coefficient of correlation between scores on it and on the whole is $\sqrt{\frac{1}{n}}$, in which $n$ is the number of times as variable the whole is as the subtest. For example, the coefficient between total scores and those on a subtest one-fourth as variable as the whole test and uncorrelated with the other three-fourths is $\sqrt{\frac{1}{4}} = +.50$. If each variable is multiplied or divided by the same factor, and it is correlated with them, the obtained coefficient is increased. For example, two series of mental ages obtained at an interval of time may be divided by chronological ages to secure intelligence quotients. If the $MA$'s and $CA$'s are positively correlated, the $IQ$'s will be more closely correlated than were the $MA$'s.

Another caution is that coefficients of correlation computed from groups of unequal variabilities, or *ranges of talent*, in the traits concerned are not directly comparable. The greater the variability, the greater will be the obtained value of the coefficient, other things being equal. For example, an obtained coefficient between chronological age and mental age in a single school grade is often rather low, whereas a similar coefficient for a whole elementary school of six or more grades is usually high. Similarly the correlation between two duplicate forms of a test obtained by giving it to a single grade is generally much less than that for the same test given to several grades. Another way of stating the same fact is that a coefficient obtained by combining groups with different means is generally larger than that in a single one of the groups.

To render coefficients obtained from different ranges comparable, formulas to estimate what coefficients would be in different ranges are available. The most frequently useful is that for two series of measures of the same thing, such as scores on duplicate forms of the same test. If subscripts **1** and **2** are employed to denote the two situations,

$$\text{estimated } r_1 = 1 - \frac{\sigma_2^2}{\sigma_1^2}(1 - r_2).$$

To illustrate this, we may assume that a coefficient of correlation of +.60 between the two forms of a test has been found in a relatively narrow range and one of +.80 for another test in a broader range, that the standard deviations are 10 and 16 points,[2] respectively, and that we wish to compare them. The estimated coefficient that would be obtained for the first test in a group with a standard deviation of 16 instead of 10 is found to be

$$1 - \frac{10^2}{16^2}(1 - .60) = + .84.$$

Since this exceeds +.80, the first test appears to possess greater self-correlation than the second. The same conclusion can also be reached by estimating the coefficient of the second test if it is given to a group with a standard deviation of only 10. In this instance the formula gives

$$1 - \frac{16^2}{10^2}(1 - .80) = + .49,$$

which is less than +.60, and therefore indicates that the second test has the lower self-correlation. In comparing coefficients of reliability of tests, the procedure just described should always be followed unless it is known that they have been derived from groups with equal variabilities.

For cases in which the two variables correlated are not similar measures of the same thing, that is, in which reliability is not involved, a more general formula must be employed. The exact one is rather complicated, but a simpler one usually yields approximations close enough to be reasonably satisfactory. It is

$$\text{estimated } r_{x_1y_1} = \frac{1}{\sigma_{x_1}}\sqrt{\sigma_{x_1}^2 - \sigma_{x_2}^2(1 - r_{x_2y_2}^2)},$$

or the same with the subscript $y$ instead of $x$ used with each $\sigma$. If the formula is employed with both $x$ and $y$ and the results averaged, a better estimate than from one alone is secured.

To illustrate the formula just given, the following measures

[2] In a real situation such as this, the standard deviations of the two duplicate forms of each test would probably be slightly different. If they are, their average should be used in the formula.

may be employed, with subscript "1" referring to the smaller range and "2" to the greater:

$$r_{x_1y_1} = +.45, \quad r_{x_2y_2} = +.77, \quad \sigma_{x_1} = 4.6, \quad \sigma_{x_2} = 6.4, \quad \sigma_{y_1} = 6.0,$$
$$\text{and } \sigma_{y_2} = 7.8.$$

The question to be answered is which value of $r$ indicates closer relationship?

As is possible for reliability data, so for these the value of either coefficient corresponding to the obtained value of the other may be estimated. The estimated value for the narrower range, estimated from that obtained for the broader by using the standard deviations of $X$, is

$$r_{x_1y_1} = \frac{1.}{4.6}\sqrt{4.6^2 - 6.4^2(1. - .77^2)} = +.46$$

and that by using those of $Y$ is

$$r_{x_1y_1} = \frac{1.}{6.0}\sqrt{6.0^2 - 7.8^2(1. - .77^2)} = +.56.$$

The average of $+.46$ and $+.56$ is $+.51$, which is the best estimate of the value corresponding to $+.77$ when the range is reduced. Since this is greater than $+.45$, the value of $r$ actually obtained for the reduced range, it indicates that a closer relationship is shown by $r = +.77$ when $\sigma_x = 6.4$ and $\sigma_y = 7.8$ than by $r = +.45$ when $\sigma_x = 4.6$ and $\sigma_y = 6.0$.

The same conclusion is reached by estimating the value of $r$ for the broader range. Thus, using the standard deviations of $X$,

$$r_{x_2y_2} = \frac{1.}{6.4}\sqrt{6.4^2 - 4.6^2(1. - .45^2)} = +.77$$

and, using those of $Y$, it equals

$$\frac{1.}{7.8}\sqrt{7.8^2 - 6.0^2(1. - .45^2)} = +.73.$$

These average $+.75$, which is less than $+.77$, and hence supports the conclusion stated above.

## Interpreting coefficients of correlation

The interpretation of coefficients of correlation in terms of how much relationship they indicate is difficult. Because of

the close connection of regression equations and certain error formulas therewith, it seems well to postpone the major treatment of the matter until after the discussion of those topics. Two or three points will be stated here, however. Several writers have suggested interpreting them in terms of adjectives such as "negligible," "low," "fair," "good," "high," and so forth, but the present writer does not believe this practice has much to recommend it. There are so many bases of judgment and comparison that the subjective element present in such terms is quite large.

Without any implication that either variable causes the other, and on the assumption, rarely completely valid, that the factors are not correlated with one another and that those common to the variables exert the same average effect as those not common,

$$r = \frac{n_c}{\sqrt{n_1 n_2}}.$$

In this $n_c$ stands for the number of common factors, $n_1$ for the number in one variable, and $n_2$ for that in the other. Thus, if, within the conditions stated, 4 factors are common to two variables, one of which has 8 and the other 10 factors, their

$$r = \frac{4}{\sqrt{8 \times 10}} = +.45;$$

if 6 are common to two, of which one has only 6 and the other 12, their

$$r = \frac{6}{\sqrt{6 \times 12}} = +.71;$$

and similarly for others.

Sometimes it is helpful to interpret a coefficient in terms of the displacement associated with it, that is, of the differences in positions of the measures in one series from those of the corresponding measures in the other. Such displacement may be stated in at least two ways. One is to divide each series into parts, generally equal, such as halves, thirds, fifths, and so on, and state to what extent corresponding measures do or do not fall in the same parts of the two series. The other is to express

the agreement and shift in terms of some measure of variability, by stating the fractions of cases whose positions in the two series differ by less than, as much as, or more than various numbers of standard, median, mean, or quartile deviations.

Since the second of the two general methods just mentioned is more commonly employed than the first, it will be amplified. By dividing $\sqrt{2 - 2r}$ into values of the measure of deviation employed .5 less[3] and .5 more than the difference in position concerned, using the results as entries in the first column of such a table of the normal curve as is given in the Appendix, finding the corresponding areas between pairs of ordinates at the given deviation distances, and finally doubling them, the per cents of cases falling within divisions the positions of which differ by the given deviation distances may be found. This sounds more difficult than it is. Let us assume that we wish to express the amounts of displacement in terms of standard deviations. If $r = +.60$, $\sqrt{2 - 2r} = .8944$; $0 \div .8944 = .0000$ and $.50 \div .8944 = .5590$. Taking these as $\sigma$ values in the first column of the standard deviation table in the Appendix, the corresponding areas between ordinates at those distances from the mean and the mean are .0000 and .2118. Twice the difference, .2118, is .4236, the fraction of cases that, for $r = +.60$, fall within the same division, $1.00\sigma$ wide, in the two series. To find the fraction falling in divisions $1.00\sigma$ apart, values of $\sigma$ of .50 and 1.50 are used. The area within $.50 \div .8944$ has already been found to be .2118. $1.50 \div .8944 = 1.6771$, and the included area corresponding to this is .4531. $2(.4531 - .2118) = .4826$, the desired fraction. Similarly it can be found that .0886 of the cases fall within divisions $2.00\sigma$ apart, .0051 within those $3.00\sigma$ apart, and less than .0001 within those more than $3.00\sigma$ apart. If we desire to know the fraction not more than a certain number of standard deviations apart, we merely total those up to and including that distance. Thus $.4236 + .4826 + .0886 = .9948$, which is the fraction not more than $2.00\sigma$ apart when $r = .60$.

To enable the reader to apply this method without going

[3] In the case of corresponding divisions, rather than those that differ in positions, a value of 0 rather than of $-.5$ should be used.

through the computation just illustrated, Table XXIII has been prepared. It is in terms of the median or quartile rather than the standard deviation and shows, for each of a number of values of the coefficient of correlation, the per cents of cases the positions of which in one series differ by the given amounts from their positions in the other. Thus, the row for $r = +.80$, for

TABLE XXIII

PER CENTS OF CASES DISPLACED IN TERMS OF MEDIAN OR QUARTILE DEVIATIONS CORRESPONDING TO GIVEN VALUES OF THE COEFFICIENT OF CORRELATION

| $r$ | Median or Quartile Deviations | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| +1.00 | 100. | | | | | | | |
| +.99 | 98. | 1.7 | | | | | | |
| +.98 | 91. | 9.2 | | | | | | |
| +.97 | 83. | 17. | | | | | | |
| +.96 | 77. | 23. | .03 | | | | | |
| +.95 | 71. | 29. | .14 | | | | | |
| +.90 | 55. | 43. | 2.4 | .02 | | | | |
| +.80 | 41. | 48. | 10. | .78 | .02 | | | |
| +.70 | 34. | 47. | 16. | 2.7 | .21 | | | |
| +.60 | 29. | 45. | 20. | 5.1 | .76 | .07 | | |
| +.50 | 26. | 43. | 22. | 7.4 | 1.6 | .22 | .02 | |
| +.40 | 24. | 40. | 23. | 9.3 | 2.5 | .49 | .07 | |
| +.30 | 23. | 38. | 24. | 11. | 3.6 | .80 | .15 | .02 |
| +.20 | 21. | 37. | 24. | 12. | 4.6 | 1.3 | .29 | .05 |
| +.10 | 20. | 35. | 24. | 13. | 5.5 | 1.8 | .46 | .09 |
| .00 | 19. | 34. | 24. | 14. | 6.3 | 2.3 | .67 | .16 |

example, shows that for this value 41 per cent of the cases fall within corresponding divisions 1.00*MdD* or *Q* wide in the two series; 48 per cent fall in divisions so wide and 1.00*MdD* or *Q* apart; 10 per cent in such divisions 2.00*MdD* or *Q* apart; .78 per cent in such divisions 3.00*MdD* or *Q* apart; and .02 per cent in such divisions 4.00*MdD* or *Q* apart. Furthermore, by adding entries from the left, one can find that 89 per cent fall in corresponding divisions not more than 1.00*MdD* or *Q* apart; 99 per cent in divisions not more than 2.00*MdD* or *Q* apart; and so on.

A more detailed and specific method of interpreting the significance of coefficients of correlation that is especially helpful

when prediction is involved requires the use of tables too elaborate to be included here. Such tables may be given in various forms. Jackson and Phillips[4] have prepared two sets of tables which will be found quite helpful for this purpose. Those of the first group show the theoretical per cent of each tenth of one variable associated with each tenth of the other for various values of the coefficient. Those of the second contain the per cents of success and failure by tenths for failure ratios of 20, 30, 40, 50, 60, 70, and 80 per cent for the same values of $r$. Both are based upon normal distributions of the two variables.

If, as is often necessary or at least convenient, a coefficient of correlation obtained from a small number of cases is to be taken as the most probable value for a large number of similar cases, of which the small number is a sample, a corrected value is likely to be more accurate than the actually obtained one. It may be obtained by the formula,

$$r_{corr} = \sqrt{1 - (1 - r^2)\left(\frac{N-1}{N-2}\right)}.$$

For example, if $r$ from 20 cases is $+.50$, the most probable value for the total population is

$$\sqrt{1 - (1 - .50^2)\left(\frac{20-1}{20-2}\right)} = +.46.$$

The effect is always, as here, to reduce the obtained value.

A probably better but more difficult formula for the same purpose is

$$r_{corr} = r - \frac{r(1 - r^2)}{2(N-1)}\left(1 - \frac{1 - 5r^2}{4(N-1)}\right).$$

For the same data this gives

$$r_{corr} = .50 - \frac{.50(1 - .50^2)}{2(20-1)}\left(1 - \frac{1 - 5 \times .50^2}{4(20-1)}\right) = +.49,$$

also smaller than uncorrected $r$ but not so much so.

[4] Jackson, Robert W. B., and Phillips, Alexander J., *Prediction Efficiencies by Deciles for Various Degree of Relationship.* Educational Research Series No. 11. Toronto: Department of Educational Research, Ontario College of Education, University of Toronto, 1945. 18 pp.

## EXERCISES AND PROBLEMS

**1.** Compute the coefficients of correlation of the following paired series:

| (a) | | (b) | | (c) | | (d) | | (e) | | (f) | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 36 | 38 | 215 | 17 | 59 | 61 | 2 | 15 | 105 | 1 | 8 | 1250 |
| 46 | 49 | 180 | 14 | 46 | 73 | 7 | 24 | 96 | 4 | 11 | 1350 |
| 96 | 91 | 169 | 9 | 55 | 72 | 3 | 11 | 121 | 0 | 4 | 1100 |
| 17 | 22 | 202 | 11 | 37 | 81 | 4 | 17 | 104 | 1 | 10 | 1400 |
| 94 | 95 | 197 | 12 | 32 | 79 | 6 | 12 | 89 | 6 | 6 | 1200 |
| 62 | 74 | 223 | 15 | 53 | 64 | 11 | 22 | 102 | 3 | 5 | 1150 |
| 31 | 28 | 186 | 13 | 47 | 72 | 7 | 13 | 109 | 2 | 24 | 1450 |
| 60 | 59 | 174 | 10 | 42 | 68 | 8 | 18 | 93 | 5 | 2 | 1050 |
| 90 | 93 | 200 | 10 | 54 | 66 | 9 | 23 | 81 | 9 | 6 | 1300 |
| 32 | 27 | 210 | 12 | 61 | 54 | 5 | 14 | 106 | 4 | 7 | 1250 |
| 58 | 66 | 196 | 8 | 47 | 75 | 4 | 25 | 112 | 1 | 13 | 1400 |
| 78 | 83 | 179 | 7 | 42 | 80 | 6 | 19 | 98 | 5 | 6 | 1250 |
| 44 | 45 | 188 | 9 | 51 | 70 | 10 | 18 | 90 | 6 | 7 | 1200 |
| 28 | 32 | 193 | 11 | 60 | 61 | 8 | 12 | 85 | 7 | 3 | 1100 |
| 72 | 71 | 205 | 14 | 52 | 74 | 3 | 29 | 123 | 1 | 1 | 1000 |
| 80 | 84 | 190 | 13 | 39 | 83 | 7 | 20 | 108 | 2 | 2 | 1050 |
| 86 | 87 | 177 | 12 | 45 | 74 | 6 | 27 | 87 | 8 | 9 | 1350 |
| 50 | 54 | 201 | 11 | 48 | 63 | 6 | 17 | 102 | 3 | 21 | 1400 |
| 79 | 82 | 209 | 15 | 53 | 68 | 12 | 16 | 99 | 4 | 7 | 1200 |
| 65 | 62 | 186 | 9 | 43 | 79 | 7 | 17 | 114 | 0 | 9 | 1250 |
| | | 194 | 11 | 38 | 44 | 8 | 14 | 103 | 2 | 3 | 1100 |
| | | 172 | 8 | 52 | 67 | 4 | 12 | 91 | 5 | | |
| | | | | 40 | 78 | 6 | 14 | 86 | 6 | | |
| | | | | 53 | 68 | | | 136 | 0 | | |
| | | | | 36 | 75 | | | | | | |

**2.** Tabulate each of the following sets of data in a correlation table:

(a) 83–89, 89–76, 72–83, 83–96, 89–74, 89–82, 83–96, 81–68, 82–74, 89–82, 75–88, 83–74, 66–75, 77–84, 86–91, 95–88, 90–80, 85–73, 89–84, 78–76, 68–67, 96–85, 85–89, 94–71, 81–76, 88–85, 86–93, 73–70, 72–84, 82–82, 76–81, 71–84, 79–80, 68–72, 93–87, 83–85, 78–76, 74–83, 85–82, 81–89, 90–92, 91–87, 73–76, 75–82, 91–95, 80–84, 77–83, 84–87, 80–88, 77–76, 74–72, 84–88, 69–68, 76–74, 83–86, 89–89.

(b) 16–10, 12–11, 10–34, 11–12, 18–16, 11–11, 10–12, 13–13, 10–10. 18–41, 15–26, 14–12, 18–39, 20–35, 18–36, 12–11, 10–24, 13–34, 19–33, 12–16, 21–24, 10–16, 10–21, 14–22, 11–27, 19–36, 17–24, 12–13, 13–18, 12–22, 14–13, 14–21, 15–30, 10–25, 11–14, 15–33, 13–32, 23–42, 16–18, 14–29, 13–25, 18–21, 15–21, 12–31, 10–17, 16–32, 12–17, 16–36, 14–11, 15–20, 11–15, 18–33, 12–23, 17–21, 15–21, 16–37.

(c) 220–1800, 70–2200, 35–2000, 200–2500, 33–2000, 279–3150, 18–1485, 135–2700, 26–1675, 80–2000, 43–1800, 17–1300, 68–2400, 39–1600, 112–2000, 63–1500, 40–2400, 197–3500, 30–1480, 23–2300, 42–1800, 52–2450, 175–3250, 182–2500, 70–2000, 243–3500, 62–2400, 113–1500, 140–3150, 168–2500, 180–2900, 75–1500, 30–2100, 95–2200, 96–2200,

37–2000, 40–2350, 244–3900, 62–2400, 52–2400, 17–1500, 140–2500, 170–3400, 135–2400, 70–2000, 158–3000, 16–1250, 118–2310, 12–1170, 74–3000, 267–3100, 23–1900, 27–2200, 135–2200, 165–4000, 90–2700, 84–2250, 18–1575, 49–2200, 27–1620, 86–2900, 160–2000, 250–3200, 111–2100, 54–1950.

(*d*) 8–119, 25–402, 1–18, 10–220, 4–92, 6–91, 2–24, 4–44, 6–116, 9–165, 10–154, 4–52, 8–119, 24–365, 6–86, 17–306, 2–33, 6–71, 1–30, 2–20, 6–73, 2–38, 6–66, 7–87, 9–155, 6–90, 8–112, 5–60, 6–121, 4–97, 4–64, 7–114, 6–115, 10–153, 5–105, 12–255, 11–291, 7–77, 6–150, 4–48, 8–199, 6–69, 5–100, 9–145, 11–205, 16–270, 6–144, 4–46, 8–118, 4–88, 1–15, 9–116, 5–120, 6–85, 10–179, 6–96, 3–39, 5–80, 15–253, 3–63, 6–108, 5–78, 6–123.

**3.** Compute the coefficient of correlation for each of the correlation tables made for the data in Exercise 2.

**4.** Compute the intercorrelations of each of the following sets of data:

| (*a*) | | | (*b*) | | | |
|---|---|---|---|---|---|---|
| *X* | *Y* | *Z* | *1* | *2* | *3* | *4* |
| 64 | 196 | 42 | 5 | 14 | 58 | 118 |
| 52 | 174 | 34 | 9 | 27 | 79 | 135 |
| 87 | 214 | 65 | 6 | 11 | 64 | 114 |
| 79 | 218 | 58 | 1 | 6 | 38 | 96 |
| 71 | 185 | 44 | 7 | 15 | 82 | 105 |
| 80 | 195 | 64 | 12 | 29 | 95 | 148 |
| 66 | 164 | 48 | 4 | 6 | 74 | 104 |
| 55 | 152 | 37 | 8 | 12 | 83 | 121 |
| 73 | 178 | 52 | 6 | 15 | 65 | 99 |
| 84 | 290 | 67 | 3 | 10 | 43 | 88 |
| 58 | 165 | 39 | 9 | 16 | 92 | 110 |
| 73 | 188 | 53 | 2 | 4 | 51 | 93 |
| 77 | 191 | 47 | 10 | 25 | 96 | 124 |
| 82 | 194 | 62 | 8 | 12 | 88 | 106 |
| 68 | 183 | 49 | 7 | 13 | 62 | 108 |
| 80 | 201 | 56 | 4 | 7 | 43 | 93 |
| 71 | 182 | 38 | 6 | 10 | 60 | 108 |
| 69 | 176 | 45 | 8 | 15 | 65 | 117 |
| 67 | 171 | 46 | 5 | 9 | 40 | 99 |
| 62 | 164 | 43 | 3 | 7 | 43 | 94 |

**5.** Which situation, of each pair stated below, indicates the closer relationship?

(*a*) $r = +.75$, $\sigma = 5.5$, or $r = +.92$, $\sigma = 8.0$.
(*b*) $r = +.40$, $\sigma$'s $= 4.0$ and $3.8$, or $r = +.85$, $\sigma$'s $= 6.4$ and $6.6$.
(*c*) $r = +.54$, $\sigma$'s $= 3.1$ and $3.2$, or $r = +.70$, $\sigma$'s $= 4.5$ and $4.7$.
(*d*) $r_{x_1y_1} = +.62$, $r_{x_2y_2} = +.81$, $\sigma_{x_1} = 3.2$, $\sigma_{x_2} = 4.2$, $\sigma_{y_1} = 12.5$, $\sigma_{y_2} = 16.5$
(*e*) $r_{x_1y_1} = +.34$, $r_{x_2y_2} = +.96$, $\sigma_{x_1} = 4.4$, $\sigma_{x_2} = 7.8$, $\sigma_{y_1} = 7.3$, $\sigma_{y_2} = 13.6$.

6. What is the coefficient of correlation in each of the following situations, assuming equal and uncorrelated factors?

(*a*) 3 common factors, 5 in one variable, 6 in other.
(*b*) 4 common factors, 5 in one variable, 12 in other.
(*c*) 8 common factors, 10 in one variable, 14 in other.

7. What is the coefficient of correlation between the total test score and that on a part of the test uncorrelated with the other parts if the part is half as variable as the whole test? What if it is one-sixth as variable?

8. Find the most probable value of the coefficient of correlation for the total population for each of the following obtained coefficients and sizes of samples:

(*a*) $r = +.65$, $N = 25$
(*b*) $r = +.86$, $N = 30$
(*c*) $r = +.48$, $N = 42$
(*d*) $r = +.74$, $N = 28$

## REFERENCES

Ezekiel, Mordecai, *Methods of Correlation Analysis.* Second Edition. Chap. XXIII; pages 455–459. New York: John Wiley & Sons, Inc., 1941.

Fisher, R. A., *Statistical Methods for Research Workers.* Eighth Edition. Pages 168–179. Edinburgh: Oliver and Boyd, 1941.

Garrett, Henry E., *Statistics in Psychology and Education.* Second Edition. Pages 251–280, 283–284, 303–305. New York: Longmans, Green & Company, Inc., 1937.

Guilford, J. P., *Fundamental Statistics in Psychology and Education.* Pages 192–208. New York: McGraw-Hill Book Company, Inc., 1942.

Kenney, John F., *Mathematics of Statistics.* Part I, pages 159–168, 175–187. New York: D. Van Nostrand Company, Inc., 1939.

Lindquist, E. F., *A First Course in Statistics.* Revised Edition. Pages 153–175. Boston: Houghton Mifflin Company, 1930.

Mode, Elmer B., *The Elements of Statistics.* Pages 276–278, 286–289, 291–301. New York: Prentice-Hall, Inc., 1941.

Monroe, Walter S., Editor, *Encyclopedia of Educational Research.* Pages 370, 372–373. New York: The Macmillan Company, 1941.

Monroe, Walter S., and Engelhart, Max D., *The Scientific Study of Educational Problems.* Pages 86–97, 110–112, 113–118, 371–377. New York: The Macmillan Company, 1926.

Peters, Charles C., and Van Voorhis, Walter R., *Statistical Procedures and Their Mathematical Bases.* Pages 91–103, 109–110, 196–203, 208–218. New York: McGraw-Hill Book Company, Inc., 1940.

Walker, Helen M., *Elementary Statistical Methods.* Pages 222–235, 237–242. New York: Henry Holt & Company, Inc., 1943.

# CHAPTER VIII

# Regression and the Interpretation of the Coefficient of Correlation

## The meaning and function of regression

It frequently occurs in educational and other work that it is desirable, even essential, to estimate or predict the score on one variable that is most likely to be associated with a given score on another variable with which the first is correlated. When this is done, there is a tendency for the predicted score to be relatively nearer the mean of its series than the score from which it was predicted is to the mean of its series. This tendency is called *regression.* The equation employed in making the prediction is the *regression equation,* and the graphic representation of this equation is the *regression curve* or *curve of regression.*

Just as correlation may be either rectilinear or curvilinear, so may regression. It also resembles correlation in that the one most frequent—indeed, almost the only—method employed is rectilinear. If the relationship between the variables is curvilinear, rectilinear regression does not enable us to make the best estimates, but because the use of curvilinear regression is much more difficult and because in many cases where regression is employed the relationship is at least approximately rectilinear, rectilinear regression is used almost to the exclusion of curvilinear.

## The regression coefficients and equations

For every pair of correlated variables, there are two regression equations, one for estimating each variable from the other.

Since they require only the means, standard deviations, and coefficient of correlation, they are easily found. There are several variations of form in which they may be expressed, the chief difference being that either the scores themselves or their deviations from means may be employed. The first is usually more convenient. Using $X$ and $Y$ to designate the two variables, the regression equation for estimating $X$ when $Y$ is known is

$$X = r\frac{\sigma_x}{\sigma_y} Y + M_x - r\frac{\sigma_x}{\sigma_y} M_y$$

and that for $Y$ when $X$ is known is

$$Y = r\frac{\sigma_y}{\sigma_x} X + M_y - r\frac{\sigma_y}{\sigma_x} M_x.$$

The term $r\frac{\sigma}{\sigma}$ is known as the *regression coefficient* and is symbolized by $b$. Hence

$$b_x = r\frac{\sigma_x}{\sigma_y} \text{ and } b_y = r\frac{\sigma_y}{\sigma_x}.$$

The regression coefficient shows the slope of the *regression line,* as the rectilinear regression curve is called, or the average amount of change in one variable associated with a change of one unit in the other. Often $b$ is substituted for $r\frac{\sigma}{\sigma}$ in the formulas above, which then become

$$X = b_xY + M_x - b_xM_y$$

and

$$Y = b_yX + M_y - b_yM_x.$$

The data in Tables XV, XVI, and XVII may be employed to illustrate the computation of regression equations. For them, $M_x = 104.25$, $M_y = 78.80$, $\sigma_x = 14.83$, $\sigma_y = 9.98$, and $r = +.84$. Substituting in the two regression equations gives

$$X = .84\,\frac{14.83}{9.98}\,Y + 104.25 - .84\,\frac{14.83}{9.98}\,78.80$$

and

$$Y = .84\,\frac{9.98}{14.83}\,X + 78.80 - .84\,\frac{9.98}{14.83}\,104.25.$$

By performing the indicated multiplication and division and combining the last two terms, these reduce to $X = 1.25Y + 5.9$ and $Y = .57X + 19.9$. Thus, if $X = 110$, for example, the value of $Y$ most likely to be associated with it is

$$.57 \times 110. + 19.9 = 83;$$

if $Y = 60$, the most likely value of $X$ associated with it is

$$1.25 \times 60. + 5.9 = 81;$$

and similarly for other cases. In these equations, the regression coefficient of $X$ on $Y$—that is, $b_x$—is 1.25 and that of $Y$ on $X$—that is, $b_y$—is .57.

In order to make the meaning of regression coefficients and equations more real, Figure 19 has been inserted. It is a cor-

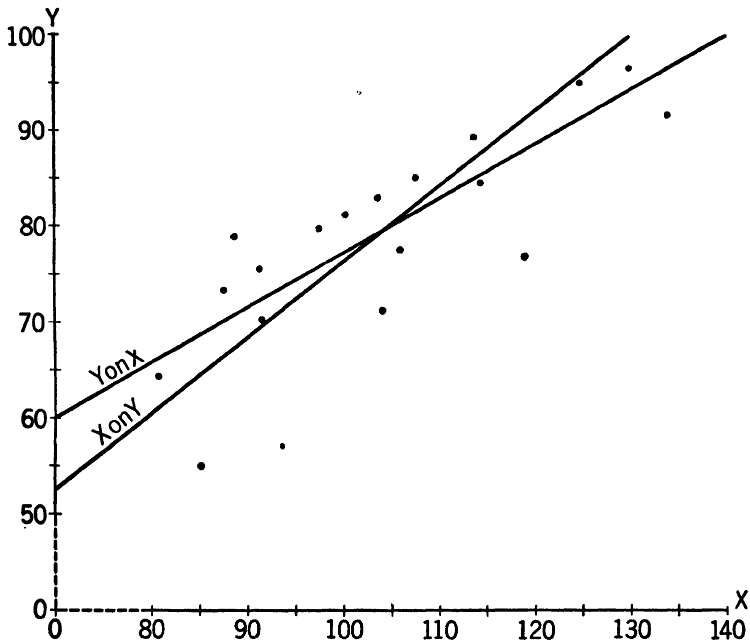



Fig. 19. Correlation Graph with Regression Lines

relation graph of the data just employed to compute the regression equations, with the individual cases and the regression lines shown. The line that is lower at the left and higher at

the right represents the regression of $X$ on $Y$, that is, the equation for estimating $X$ when $Y$ is known, and the one that is higher at the left and lower at the right that of $Y$ on $X$. It will be observed that the two lines intersect at the means.

The horizontal distance, parallel to the $X$-axis, of each point from the $X$ on $Y$ line represents the difference between the estimated and the actual value of each $X$ case, and the vertical distance, parallel to the $Y$-axis, from the $Y$ on $X$ line represents the difference between the estimated and actual value of each $Y$ case. For example, the equation for $Y$ gives $.57 \times 135. + 19.9 = 97$ as the most likely value of $Y$ associated with $X = 135$, whereas it actually is 92. The difference, 5, is what is known as the *error of estimate*. The lower the correlation, the larger the errors; and the greater, either positive or negative the correlation, the smaller the errors; until when $r = \pm 1.00$, all the cases lie on the line of regression and there are no errors. The size of errors in scores estimated by regression equations will be discussed more fully in a later section.

As a second illustration of the computation of regression equations, the data in Tables XVIII, XIX, XX, and XXI will be used. For them, $M_x = 246.4$, $M_y = 106.8$, $\sigma_x = 57.78$, $\sigma_y = 14.04$, and $r = +.89$. Therefore,

$$b_x = .89 \frac{57.78}{14.04} = 3.66$$

and

$$b_y = .89 \frac{14.04}{57.78} = .22;$$

and the regression equations are

$$X = 3.66Y + 246.4 - 3.66 \times 106.8$$

and

$$Y = .22X + 106.8 - .22 \times 246.4,$$

which reduce to $X = 3.66Y - 144.8$ and $Y = .22X + 53.5$.

Regression coefficients, similarly to correlation coefficients, may be computed without actually finding the means, standard

deviations, and coefficient of correlation. The formulas, in terms of actual scores, are

$$b_x = \frac{\Sigma XY - \dfrac{\Sigma X \cdot \Sigma Y}{N}}{\Sigma Y^2 - \dfrac{(\Sigma Y)^2}{N}}$$

and

$$b_y = \frac{\Sigma XY - \dfrac{\Sigma X \cdot \Sigma Y}{N}}{\Sigma X^2 - \dfrac{(\Sigma X)^2}{N}}$$

and in terms of deviations the same with $x$ and $y$ instead of $X$ and $Y$, respectively.

For the benefit of readers familiar with the method of least squares, it seems well to mention that the regression equations are the same as those of the lines of best fit employed in that method. They may be found by the solution of simultaneous equations involving terms that correspond to $b$ and the constant that is added or subtracted.

Sometimes scores are expressed in terms of their standard deviation distances from the mean. In other words, such a score equals $\frac{X - M}{\sigma}$, in which $X$ represents the obtained score. It is variously called *standard, sigma,* and *z-score.* When scores are so expressed, the standard deviation of their distribution becomes 1.00. Therefore $\beta$ (beta), the symbol employed for the regression coefficient of standard scores, equals $r\frac{1}{1}$ or just $r$. Using this in the regression equations gives, for such scores,

$$z_x = rz_y \text{ and } z_y = rz_x.$$

The terms *dependent variable* and *independent variable* often occur in connection with regression. The *dependent variable* is the one estimated or predicted from the other, whereas the latter is known as the *independent variable.*

## The interpretation of coefficients of correlation and regression equations

Generally the most helpful method of stating the degree of relationship indicated by a given value of the coefficient of correlation—which is another way of saying the amount of error present in scores estimated by a regression equation—is by means of the coefficients of alienation and of dependability. The *coefficient of alienation*, symbolized by $k$, equals $\sqrt{1 - r^2}$. The *coefficient of dependability*, symbolized by $E_m$, equals $1 - k$[1] or $1 - \sqrt{1 - r^2}$. Sometimes the latter is multiplied by 100 and called the *index of prediction*; that is,

$$I_p = 100 - 100k \text{ or } 100 - 100\sqrt{1 - r^2}.$$

To interpret $k$ and $E_m$, one should think of a scale of the error involved in estimated scores extending from no error at all—the condition when $r = \pm 1.00$—down to the amount of error involved when all are simply estimated at the mean—a condition accompanying which $r = .00$. The coefficient of alienation, $k$, shows what fraction the amount of error connected with a given value of $r$ is of the amount when all are estimated at the mean. The coefficient of dependability, $E_m$, shows the fractional reduction in amount of error for the value of $r$ with which it is associated. The second and third columns of Table XXIV give the values of $k$ and $E_m$ for each of a number of values of $r$. For example, if $r = \pm .95$, the amount of error in scores estimated by the regression equation is .3122 as great as if all were taken at the mean and, conversely, .6878 better than if all were so taken.

The most evident fact revealed by this table is that, for high values of $r$, small decreases therein are accompanied by large increases in $k$ and large decreases in $E_m$; whereas, for small values of $r$, large increases are accompanied by small changes in $k$ and $E_m$. Many persons are accustomed to think of values of $r$ in the .90's as indicating quite close relationship, whereas

[1] A more accurate value is given by

$$E_m = 1 - \left(\sqrt{\frac{N-1}{N-2}}\right) k,$$

but the formula is rarely employed.

TABLE XXIV

VALUES OF $k$, $E_m$, $g$, AND $E_g$ CORRESPONDING TO GIVEN VALUES OF $r$.

| $r$ | $k$ | $E_m$ | $g$ | $E_g$ |
|---|---|---|---|---|
| ±1.00 | .0000 | 1.0000 | .0000 | 1.0000 |
| ±.99 | .1411 | .8589 | .0998 | .9002 |
| ±.98 | .1990 | .8010 | .1407 | .8593 |
| ±.97 | .2431 | .7569 | .1719 | .8281 |
| ±.96 | .2800 | .7200 | .1980 | .8020 |
| ±.95 | .3122 | .6878 | .2208 | .7792 |
| ±.94 | .3412 | .6588 | .2413 | .7587 |
| ±.93 | .3676 | .6324 | .2599 | .7401 |
| ±.92 | .3919 | .6081 | .2771 | .7229 |
| ±.91 | .4146 | .5854 | .2932 | .7068 |
| ±.90 | .4359 | .5641 | .3082 | .6918 |
| ±.866 | .5000 | .5000 | .3536 | .6464 |
| ±.85 | .5268 | .4732 | .3725 | .6275 |
| ±.80 | .6000 | .4000 | .4243 | .5757 |
| ±.75 | .6614 | .3386 | .4677 | .5323 |
| ±.70 | .7141 | .2859 | .5050 | .4950 |
| ±.65 | .7599 | .2401 | .5373 | .4627 |
| ±.60 | .8000 | .2000 | .5657 | .4343 |
| ±.55 | .8352 | .1648 | .5906 | .4094 |
| ±.50 | .8660 | .1340 | .6123 | .3877 |
| ±.40 | .9165 | .0835 | .6481 | .3519 |
| ±.30 | .9539 | .0461 | .6745 | .3255 |
| ±.20 | .9798 | .0202 | .6928 | .3072 |
| ±.10 | .9950 | .0050 | .7036 | .2964 |
| .00 | 1.0000 | .0000 | .7071 | .2929 |

the table shows that this is far from true. To illustrate this graphically, Figure 20 has been inserted.

The entries in the column headed $k$ may also be employed somewhat differently. The common method of describing the size of the *errors of estimate*—that is, the differences between scores estimated by regression equations and actual scores—is to state a measure of their variability. Of such measures the *standard error of estimate*, which is merely the standard deviation of the errors, and the *probable error of estimate*, which is the median of the errors, are the two most commonly employed. For the former, the usual symbol is $\sigma_{1\cdot2}$ or $\sigma_{est}$, but sometimes $\epsilon_{1\cdot2}$ (epsilon) or $\epsilon_{est}$ is employed,[2] and, for the latter, $PE_{1\cdot2}$ or

[2] The use of $\epsilon$ instead of $\sigma$ is preferable, in order to avoid confusion with $\sigma$ as the standard deviation, but such use is rare.

$PE_{est}$. Since the formula for $\epsilon_{1 \cdot 2}$ is $\sigma_1\sqrt{1. - r^2}$ or $\sigma_1 k$, the entries in the second column of Table XXIV may be employed in computing $\epsilon_{1 \cdot 2}$. Likewise, since

$$PE_{1 \cdot 2} = \text{either } .6745\epsilon_{1 \cdot 2} \text{ or } MdD_1k,$$

the entries may be used for that also.

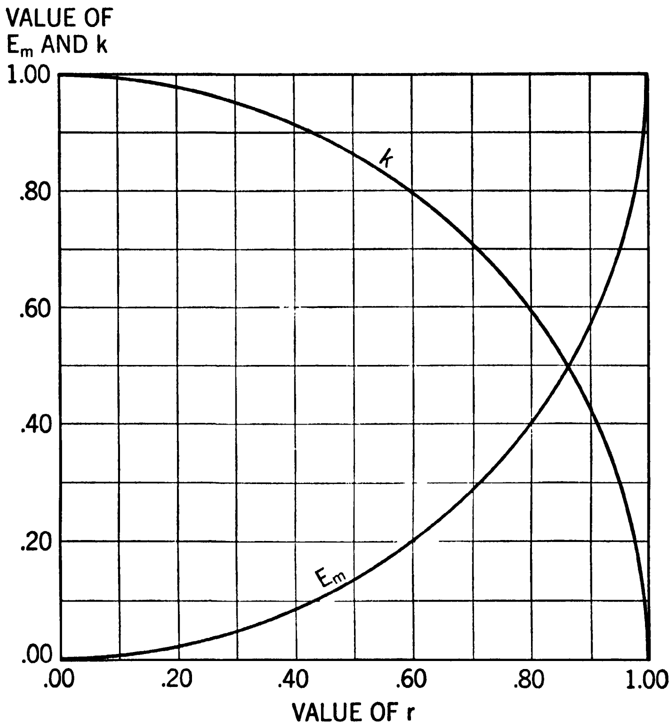



**Fig. 20. Values of $E_m$ and $k$ Corresponding to Given Values of $r$**

For the data in Tables XV, XVI, and XVII, already employed for correlation and regression,

$$\epsilon_{x \cdot y} = 14.83\sqrt{1. - .84^2} = 8.05$$

and

$$\epsilon_{y \cdot x} = 9.98\sqrt{1. - .84^2} = 5.42.$$

In other words, if the regression equation for $X$, which is

$X = 1.25Y + 5.9$, is used to estimate its values, 68.27 per cent of them will be in error by less than 8.05, and so on.

$$PE_{x \cdot y} = .6745\epsilon_{x \cdot y},$$

which is $.6745 \times 8.05$ or 5.43; hence 50 per cent of the errors will be less than 5.43; and so on. Readers should note that the first symbol in the subscript of an error of estimate denotes the variable being estimated and the second denotes the one from which it is estimated.

The standard and probable errors of estimate, being standard and median deviations, may be interpreted as any other such deviations, following the discussion in Chapter VI. That is, 68.27 per cent of the errors of estimate are less than their standard error, 95.45 per cent are less than twice their standard error, and so on, for any value of it or of the probable error, according to the tables in the Appendix. In this, as in all similar cases, these interpretations are based upon the assumption of normal distribution.

If secured from a small sample and taken as applying to the total population, the standard or probable error of estimate, just as the coefficient of correlation, should be corrected. The correction is merely the multiplication of the obtained value by

$$\sqrt{\frac{N-1}{N-2}}.$$

Thus, for the 20 cases for which $\epsilon_{x \cdot y}$ was found to be 8.05, its corrected value is given by

$$\epsilon_{x \cdot y_{corr}} = 8.05\sqrt{\frac{19}{18}} = 8.27.$$

The last two columns of Table XXIV may be used in another method of stating the amount of error or reduction of error associated with values of the coefficient of correlation. It has not been frequently employed, but is sometimes confused with the other. This method differs from that associated with $k$ and $E_m$ in that the point of maximum error is taken as being the situation resulting when the scores in two series of variables are matched by pure chance or guess, such as by drawing from

boxes or hats. The symbol $g$ is used to represent the guessing or error element; $E_g$, for that of reduction in error or certainty. For example, if $r = \pm.75$, the errors in scores estimated by the regression equation are .4677 as large as they would be in pure guesses, or the reduction in error is .5323 over that. The formula for $g$ is merely $\sqrt{\frac{1 - r^2}{2}}$ or $.7071k$, and that for $E_g$ is $1 - g$ or $1 - .7071k$ or $.7071E_m + .2929$.

Another method of interpretation involves $r^2$ and $k^2$, known as the coefficient of determination and of nondetermination, respectively. This method applies only if one variable is associated with all of another plus one or more other factors. In such a situation the *coefficient of determination*, that is, $r^2$, represents the fraction of the variance in the first variable associated with the second, and the *coefficient of nondetermination*, $k^2$, represents the fraction of the variance in the first associated with other factors than the second.[3] Thus if, in the situation from which the data employed in Tables XV, XVI, XVII, and elsewhere were obtained, all of $X$ plus other factors are associated with $Y$, $.84^2$ or 71 per cent of the factors in $Y$ are associated with $X$ and 29 per cent are not.

The coefficients of correlation and alienation and of determination and nondetermination are closely connected with the standard deviations and variances of the variables. If $\sigma_{y_1}$ is the standard deviation of the portion of $Y$ associated with $X$ and $\sigma_{y_2}$ the standard deviation of the portion of $Y$ not associated with $X$,

$$\sigma_y = \sqrt{\sigma_{y_1}^2 + \sigma_{y_2}^2},\ r = \frac{\sigma_{y_1}}{\sigma_y},\ k = \frac{\sigma_{y_2}}{\sigma_y},\ r^2 = \frac{\sigma_{y_1}^2}{\sigma_y^2},\ \text{and } k^2 = \frac{\sigma_{y_2}^2}{\sigma_y^2}.$$

If one variable—in this case, $Y$—is associated with only a part of another variable—here $X$—plus other factors, the determination of the first by the second is the square of their coefficient of correlation divided by the square of the coefficient between the common factor or factors and the second variable.

[3] Readers should recall that $r^2 + k^2 = 1.00$.

That is, using $c$ to denote the common factor, the determination of $Y$ by $X$ is

$$\frac{r_{xy}^2}{r_{cx}^2}$$

and the nondetermination is

$$1 - \frac{r_{xy}^2}{r_{cx}^2}.$$

## Estimated true scores

It is sometimes desirable to secure the best estimates of what scores would be if they were perfectly reliable. The formula for estimating true scores is

$$X'_\infty = rX + (1 - r)M \text{ or } r(X - M) + M,$$

in which $X'_\infty$ stands for the estimated true score,[4] $r$ for the coefficient of reliability—that is, of correlation between duplicate forms or two administrations of the test—and $X$ for the actually obtained or *fallible score.*

To illustrate the application of this formula, we may assume that for a certain test the mean score is 66 and the coefficient of reliability is +.88. Then the estimated true score for a fallible score of 80 is

$$.88 \times 80. + (1 - .88)66. = 78.;$$

that for one of 58 is

$$.88 \times 58. + (1 - .88)66. = 59.;$$

and so one for others. By performing the indicated operations the last term becomes 7.92; hence for this situation the formula for any estimated true score becomes $.88X + 7.92$. The reader should note that an estimated true score is always nearer the mean than is the fallible score from which it is obtained.

### EXERCISES AND PROBLEMS

**1.** Compute the regression equations of the data in Exercises 1 and 2 on pages 134 and 135.

**2.** In Exercise 1, Part (*a*), on page 134, what is the estimated value

[4] The prime (′) regularly signifies an estimated rather than an actual score and the subscript $\infty$ (infinity) refers to a true measure.

of the second variable corresponding to each of the following values of the first? 40, 56, 81, 95.

**3.** In Exercise 2, Part (*b*), on page 134, what is the estimated value of the first variable corresponding to each of the following values of the second? 14, 18, 26, 30.

**4.** Compute the values of $r^2$, $k$, $k^2$, $E_m$, $g$, and $E_g$ for each part of Exercises 1 and 2 on pages 134 and 135.

**5.** Find the standard and probable errors of estimate for each variable for the data in Exercises 1 and 2 on pages 134 and 135.

**6.** Find the corrected values of the standard and probable errors of estimate found in Exercise 5 if they are to be taken as applying to a total population.

**7.** If the first variable is considered to be completely associated with the second, find the standard deviation of that part of the second associated with the first, and of that part of the second not associated with the first, for each part of Exercises 1 and 2 on pages 134 and 135.

**8.** If the common factor in the variables has a coefficient of correlation of .70 with one of them, find the determination and nondetermination of the other by it for the data in each part of Exercises 1 and 2 on pages 134 and 135.

**9.** Find the equation for estimated true scores in each series for Parts(*a*) of Exercises 1 and 2 on page 134.

**10.** Apply the equations found in Exercise 9 to estimate the true scores for fallible scores of 45 and 78 in the first series in Part (*a*), Exercise 1, and for fallible scores of 76 and 92 in the second series in Part (*a*), Exercise 2.

## REFERENCES

Ezekiel, Mordecai, *Methods of Correlation Analysis.* Second Edition. Chaps. III–IX. New York: John Wiley & Sons, Inc., 1941.

Garrett, Henry E., *Statistics in Psychology and Education.* Second Edition. Pages 289–303, 305–309. New York: Longmans, Green & Company, Inc., 1937.

Guilford, J. P., *Fundamental Statistics in Psychology and Education.* Chap. X; pages 211–224. New York: McGraw-Hill Book Company, Inc., 1942.

Kenney, John F., *Mathematics of Statistics.* Part I, pages 168–175, 187–198; Part II, Chap. IV. New York: D. Van Nostrand Company, Inc., 1939.

Lindquist, E. F., *A First Course in Statistics.* Revised Edition. Pages 175–204. Boston: Houghton Mifflin Company, 1942.

Mode, Elmer B., *The Elements of Statistics.* Pages 278–286, 289–291. New York: Prentice-Hall, Inc., 1941.

Monroe, Walter S., Editor, *Encyclopedia of Educational Research.* Pages 594, 839–841, 937. New York: The Macmillan Company, 1941.

Monroe, Walter S., and Engelhart, Max D., *The Scientific Study of Educational Problems.* Chap. X; pages 99–100, 389–395. New York: The Macmillan Company, 1926.

Peters, Charles C., and Van Voorhis, Walter R., *Statistical Procedures and Their Mathematical Bases.* Pages 110–123. New York: McGraw-Hill Book Company, Inc., 1940.

Walker, Helen M., *Elementary Statistical Methods.* Chaps. XII, XIV; pages 234–237. New York: Henry Holt & Company, Inc., 1943.

# CHAPTER IX

# The Reliability of Scores

## The coefficient of reliability

It has already been stated that the coefficient of correlation between two administrations of the same test, or two forms thereof, is called the *coefficient of reliability.* Other synonymous terms are *coefficient of self-correlation* and *retest coefficient.* The definition just given may be broadened to include any two series of scores or measures of the same individuals obtained by the same measuring instrument or procedure. In terms of its significance, the coefficient of reliability is the fraction of the test score that measures whatever it is that the test measures, the remaining part being due to variable errors.

To distinguish the coefficient of reliability from other applications of the coefficient of correlation, its symbol often has a subscript consisting of the Arabic and Roman forms of the same figure. If a single test is in question, it is regularly $r_{1\mathrm{I}}$; for a second test it is $r_{2\mathrm{II}}$; and so on.

Although the coefficient of reliability is the most frequently employed measure of the reliability of a test or a set of scores, it is far from satisfactory. The chief reason for its weakness is the fact, already stated in Chapter VII, that it is so largely affected by the range of scores. One should not judge, therefore, the relative reliability of measuring instruments by their coefficients of reliability unless he knows that the ranges of scores upon which they are based are at least approximately equal or that they have been made comparable by the procedure given in Chapter VII. Later in this chapter other measures of test or score reliability will be presented.

## The Spearman-Brown formula

An important factor in reliability is length. It is a matter of common observation that a large sample of anything usually yields more accurate or reliable information concerning the whole than does a small sample. The same condition holds in testing. A test of 40 elements, for example, ordinarily results in more reliable scores than does one of 20 elements. The *Spearman-Brown formula,* sometimes called the *prophecy formula,* deals with the relationship between length of measuring instrument and reliability. It is based upon several assumptions, as follows: that the material in the instruments is homogeneous, that is, similar in type of content, difficulty, form, time per element, and other pertinent characteristics; that the conditions of testing are the same; and that fatigue does not enter in sufficiently to affect performance on the longer test. When these conditions are met, the formula gives results in close agreement with empirical data. It is

$$r_n = \frac{nr_{11}}{1 + (n - 1)r_{11}},$$

in which $n$ refers to the number of times a test is lengthened or the ratio of length of the longer to the shorter and $r_n$ to the coefficient of reliability of the longer test.

As an example of the use of this formula, if a test of 20 elements for which $r_{11} = +.50$ is lengthened to have 60 elements, that is, to be three times as long, and the conditions stated above are met, the coefficient of reliability of the lengthened test may be expected to be close to

$$\frac{3 \times .50}{1 + (3 - 1).50} = +.75.$$

As a second example, we may determine whether a test of 30 elements with $r_{11} = +.45$ would be more or less reliable than one of 100 elements with $r_{11} = +.72$ if it were made the same length. Since $100 \div 30 = 3\frac{1}{3}$, the formula gives

$$r_{n(\text{here } 3\,1/3)} = \frac{3\frac{1}{3} \times .45}{1 + (3\frac{1}{3} - 1).45} = +.73,$$

which indicates that it would be slightly more reliable.

The most frequent application of this formula is in determining the coefficient of reliability of measuring instruments which have only one form or for which the administration of two forms is impracticable. The procedure is to compute two scores for each individual, usually one from the odd-numbered and one from the even-numbered elements. The coefficient of correlation between these two series of scores is the coefficient of reliability of half the test. To get that of the whole test, the Spearman-Brown formula, with $n = 2$, is employed. This is

$$\frac{2r_{\mathrm{II}}}{1 + (2 - 1)r_{\mathrm{II}}}$$

which reduces to

$$\frac{2r_{\mathrm{II}}}{1 + r_{\mathrm{II}}}.$$

Thus, if the coefficient of reliability of half a test is +.65, that of the whole test is

$$\frac{2 \times .65}{1.65} = +.79.$$

Since the conditions surrounding one giving are practically identical for scores on odds and on evens, whereas those for two administrations are liable to differ, coefficients of reliability secured thus are likely to be slightly higher than those from two administrations of either the same or duplicate forms. There are also other reasons why the two types of coefficients are not exactly comparable, but they are commonly used as if they were.

Since the Spearman-Brown formula applies to scores on series of homogeneous tests administered under similar conditions, the reliability of total scores derived therefrom may be estimated by it. The coefficients of reliability of the tests composing the total will vary, but the average coefficient of those of a given length may be used for $r_{\mathrm{II}}$. For example, if the average coefficient of 20 short tests of the same length is +.40, that of the total score may be expected to approximate

$$\frac{20 \times .40}{1 + (20 - 1).40} = +.93.$$

It is sometimes useful to reverse the formula, to find how much to increase length to secure a desired higher reliability coefficient. For example, if a test of 50 elements has a coefficient of +.63, how many elements are necessary to yield a coefficient of +.90? Solving the formula for $n$ gives

$$n = \frac{r_n(1 - r_{11})}{r_{11}(1 - r_n)}.$$

Substituting in this gives

$$n = \frac{.90(1 - .63)}{.63(1 - .90)} = 5^2/_7;$$

$5^2/_7 \times 50 = 264$ elements, the number needed. The same formula may be used in reducing length and reliability. Thus, if a test of 300 elements has a coefficient of +.97 and one of +.95 is sufficient,

$$n = \frac{.95(1 - .97)}{.97(1 - .95)} = .5876 \text{ and } .5876 \times 300 = 176,$$

the needed number.

The Spearman-Brown formula also applies approximately in other situations, such as those involving the number of judges or raters and of suggested responses in multiple-answer tests. If elements in such a test are of approximately equal difficulty, or if scores are weighted according to difficulty, the relationship between reliability and number of responses follows this formula.

## Attenuation and its correction

The presence of variable errors in correlated measures tends to reduce obtained coefficients of correlation below what they would be if no such errors were present. This reduction in coefficients of correlation is called *attenuation*. By means of the formula to be given it can be eliminated, and the value which a coefficient would have if both series were perfectly reliable can be estimated. In most instances there is no good reason for doing so, but in some experimental and other research studies, particularly some of those that endeavor to determine causation, it is helpful to do so.

The general form of the formula to be used to correct for attenuation involves dividing a measure of the correlation between the two series by a measure of their reliability. In order that the latter be available, there must be two series of measures of each variable. Therefore, there are, in such cases, four series in all. The two series of measures of one variable may be symbolized by 1 and I, those of the other by 2 and II. Four correlation coefficients between the two variables are possible, since 1 may be correlated with both 2 and II, and 2 with both 1 and I also. The two coefficients of reliability are, of course, 1 with I and 2 with II. When more than one coefficient is employed in either numerator or denominator of the formula, the geometric mean is used as being the best average thereof.

The best form of the formula contains all possible coefficients; thus,

$$r_{12_{corr}} = \frac{\sqrt[4]{r_{12}r_{1\mathrm{II}}r_{\mathrm{I}2}r_{\mathrm{III}}}}{\sqrt{r_{1\mathrm{I}}r_{2\mathrm{II}}}}.$$

Since almost as good results are secured by using only two coefficients, 1 with II and I with 2, in the numerator, the usual form employed is

$$\sqrt{\frac{r_{1\mathrm{II}}r_{\mathrm{I}2}}{r_{1\mathrm{I}}r_{2\mathrm{II}}}}.$$

It is possible to use a single coefficient, without the radical sign, in the numerator, but this is not recommended.

To illustrate correction for attenuation by the usual formula, let us suppose that the coefficient of 1 with II is +.56; of I with 2, +.61; the reliability coefficient of 1 with I, +.78; and of 2 with II, +.91. Then,

$$r_{12_{corr}} = \sqrt{\frac{.56 \times .61}{.78 \times .91}} = +\ .69,$$

the estimated coefficient if both variables were perfectly reliable.

Inspection of the formula makes it evident that the higher the reliabilities of the two series of scores, the less the increase in the coefficient between them when the effect of unreliability is removed, and vice versa. If they are perfectly reliable, there is no increase at all.

## The index of reliability

The *index of reliability* is the coefficient of correlation between the scores actually obtained and the theoretical true scores. It is the square root of the coefficient of reliability, that is, $\sqrt{r_{11}}$; therefore, except when both are .00 or 1.00, it is always larger than the coefficient. The index really is a better measure of reliability than the coefficient, since agreement with true scores is a better criterion of consistency than is agreement with another series of fallible scores, but it has never come into common use. Since the index and coefficient are so closely and unvaryingly related to each other, both lead to the same conclusions in the comparison of measuring instruments as to their reliability. Likewise the index is affected by the range of talent or variability of scores; hence, it is subject to the same limitations of interpretation in this respect as is the coefficient.

## Errors of measurement

It has already been stated that other measures of the reliability of scores are preferable to the coefficient. The error of measurement, and its ratios to the mean and standard deviation, are the measures meant. An *error of measurement,* also called *error of a score* and *error of observation,* is the difference between an actual and a true score. In other words, it is the amount by which an obtained score differs from what it would be if it were perfectly accurate. Another definition of an error of measurement is that it is the error of estimate involved with an actual score and a perfectly reliable true score.

Errors of measurement, just as errors of estimate and others, are generally measured by their standard and median deviations, called the *standard error of measurement* and the *probable error of measurement.* The usual formula[1] for the first is

[1] Several other formulas are available. One that has much merit but is seldom used is

$$\epsilon_{1\cdot\infty} = \sqrt{\frac{\Sigma(1 - \mathrm{I})^2 - \dfrac{(\Sigma 1 - \Sigma \mathrm{I})^2}{N}}{2N}}.$$

Another is

$$\sqrt{\frac{\Sigma d^2}{N}},$$

in which $d$ is the difference between scores on equivalent halves.

$$\epsilon_{1\cdot\infty} \text{ or } \sigma_{1\cdot\infty} \text{ or } \epsilon_{meas} \text{ or } \sigma_{meas} = \sigma\sqrt{1 - r_{1I}}$$

and for the second

$$PE_{1\cdot\infty} \text{ or } PE_{meas} = .6745\sigma\sqrt{1 - r_{1I}} \text{ or } MdD\sqrt{1 - r_{1I}}.$$

These formulas are obtained from those for the standard and probable errors of estimate by substituting the index of reliability for the coefficient of reliability employed therein. In each case the coefficient used is that of the correlation between the two series of scores involved. Thus,

$$\epsilon_{1\cdot\infty} = \sigma\sqrt{1 - (\sqrt{r_{1I}})^2},$$

which equals $\sigma\sqrt{1 - r_{1I}}$. Since reliability usually involves two forms of a test and their standard or median deviations are likely to differ slightly, the average value is regularly taken. Hence, the formula for the standard error of measurement becomes

$$\frac{\sigma_1 + \sigma_I}{2}\sqrt{1 - r_{1I}}$$

and that for the probable error

$$.6745\,\frac{\sigma_1 + \sigma_I}{2}\sqrt{1 - r_{1I}}$$

or

$$\frac{MdD_1 + MdD_I}{2}\sqrt{1 - r_{1I}}.$$

Thus if a test has a coefficient of reliability of +.84 and the standard deviations of its two forms are 10.2 and 10.4, its

$$\epsilon_{1\cdot\infty} = \frac{10.2 + 10.4}{2}\sqrt{1 - .84} = 4.1$$

and its

$$PE_{1\cdot\infty} = .6745 \times 4.1 = 2.8.$$

The standard and probable errors of measurement are interpreted as are any standard and median deviations and constitute satisfactory measures of the absolute sizes of test errors due to unreliability. Because they are absolute rather than relative, however, they are not valid for comparing tests unless similar scores thereon happen to have the same significance.

Just as an error of a foot is generally a serious one in the measurement of the length of a small room but not in the measurement of the distance from one town to another, so an error of 5 points is much more serious on a test of 25 elements than on one of 250 elements. To secure relative measures of error, absolute measures thereof must be compared with measures of the opportunity for error.

Two measures of opportunity for error have been commonly employed for the purpose just stated. They are the mean and the standard deviation. Either the standard or the probable error of measurement may be compared with them by simple division, it being used as the dividend and the mean or the standard deviation as the divisor. The probable error is employed in this manner more frequently than is the standard error. If two forms of a test with different means and standard deviations are concerned, the average values are used. Therefore, the usual forms of the formulas for the ratios are

$$.6745\,\frac{\frac{\sigma_1 + \sigma_{\mathrm{I}}}{2}\sqrt{1 - r_{1\mathrm{I}}}}{\frac{M_1 + M_{\mathrm{I}}}{2}},$$

which reduces to

$$.6745\,\frac{(\sigma_1 + \sigma_{\mathrm{I}})\sqrt{1 - r_{1\mathrm{I}}}}{M_1 + M_{\mathrm{I}}},$$

and

$$.6745\,\frac{\frac{\sigma_1 + \sigma_{\mathrm{I}}}{2}\sqrt{1 - r_{1\mathrm{I}}}}{\frac{\sigma_1 + \sigma_{\mathrm{I}}}{2}},$$

which becomes

$$.6745\sqrt{1 - r_{1\mathrm{I}}}.$$

If the means of the two forms of the test mentioned on the previous page are 68.3 and 69.1, the ratios of its probable error of measurement to the mean and standard deviation are

$$\frac{2.8}{\frac{68.3 + 69.1}{2}} = .04,$$

and

$$\frac{2.8}{\frac{10.2 + 10.4}{2}} \text{ or } .6745\sqrt{1. - .84} = .27.$$

Each of these two relative measures has its merits and demerits, hence its adherents and opponents. The mean, and therefore the first ratio, can have its value materially changed merely by adding a number of very easy elements to those already in a test, by removing from a test its easiest elements, and in various other ways that have little or no effect upon the real reliability of the scores that result from it, in other words, upon the size of the errors. Another way of stating the same point is that, since most tests do not have true zero points, their true means are not so exactly determined as appears on the surface. On the other hand, the second ratio depends directly upon the coefficient of reliability, hence upon the range of talent, which we have seen has a disturbing effect. The writer believes that it is advisable to make use of both of the suggested ratios, but to give somewhat more weight to the first, that is, the one based upon the mean, than to the second. If they agree in indicating that one test is more reliable than another, that conclusion may be accepted with a high degree of confidence.

## EXERCISES AND PROBLEMS

**1.** Compute the coefficients of reliability, the indices of reliability, the standard and probable errors of measurement, and the ratios of the latter to the means and standard deviations for the data in Parts (*a*) of Exercises 1 and 2 on page 134.

**2.** Find the coefficients of reliability of tests with original coefficients and changes in lengths as follows:

| | (*a*) | (*b*) | (*c*) | (*d*) | (*e*) |
|---|---|---|---|---|---|
| $r_{\text{II}}$ | +.66 | +.35 | +.96 | +.77 | +.81 |
| Times as long | 2. | 8. | .5 | 3. | 2.5 |

**3.** How many times as long must tests with the coefficients of reliability given below be made to have the stated desired coefficients?

| | (*a*) | (*b*) | (*c*) | (*d*) | (*e*) |
|---|---|---|---|---|---|
| Actual coefficient | +.60 | +.86 | +.98 | +.42 | +.73 |
| Desired coefficient | +.90 | +.97 | +.95 | +.87 | +.92 |

**4.** Find the coefficient of correlation corrected for attenuation in each of the following situations:

(*a*) $r_{1\mathrm{II}} = +.38$, $r_{\mathrm{I}2} = +.42$, $r_{1\mathrm{I}} = +.85$, $r_{2\mathrm{II}} = +.67$.
(*b*) $r_{12} = +.45$, $r_{1\mathrm{I}} = +.92$, $r_{2\mathrm{II}} = +.86$.
(*c*) $r_{12} = +.63$, $r_{1\mathrm{II}} = +.65$, $r_{\mathrm{I}2} = +.58$, $r_{1\mathrm{II}} = +.66$, $r_{1\mathrm{I}} = +.82$, $r_{2\mathrm{II}} = +.84$.
(*d*) $r_{1\mathrm{II}} = +.79$, $r_{\mathrm{I}2} = +.81$, $r_{1\mathrm{I}} = +.89$, $r_{2\mathrm{II}} = +1.00$.

## REFERENCES

Garrett, Henry E., *Statistics in Psychology and Education.* Second Edition. Chap. XII; pages 311–324, 333–335. New York: Longmans, Green & Company, Inc., 1937.

Guilford, J. P., *Fundamental Statistics in Psychology and Education.* Pages 273–284. New York: McGraw-Hill Book Company, Inc., 1942.

Lindquist, E. F., *A First Course in Statistics.* Revised Edition. Pages 216–227. Boston: Houghton Mifflin Company, 1942.

Lindquist, E. F., *Statistical Analysis in Educational Research.* Pages 233–235. Boston: Houghton Mifflin Company, 1940.

Monroe, Walter S., and Engelhart, Max D., *The Scientific Study of Educational Problems.* Pages 130–136, 149–152, 199–209. New York: The Macmillan Company, 1926.

Peters, Charles C., and Van Voorhis, Walter R., *Statistical Procedures and Their Mathematical Bases.* Pages 191–195, 203–208. New York: McGraw-Hill Book Company, Inc., 1940.

# CHAPTER X

# Rank Correlation

## Nature and function

Rank correlation has been suggested as an easier means of measuring and expressing the relationship between two variables than is product-moment correlation. It is a method, or rather includes two methods, that takes account only of ranks or relative positions of scores, not of their exact values. Since it can be used only with measures in columns, not with those in correlation tables, since its computation rapidly becomes more laborious as the number of cases increases, since it is less reliable than product-moment correlation, also less valid for data approximating normality and of various other shapes, it is rarely employed with more than thirty or forty cases. In some instances involving larger numbers of cases only ranks are available; hence product-moment correlation cannot be found, and so rank correlation may be employed. If there are tied ranks in one or both series, the resulting value of rank correlation is too large. It can be, but rarely is, corrected by a little-known formula.

In general, rank correlation may be considered an easy, approximate method to be employed when the number of cases is so small that the slight difference in reliability between it and product-moment correlation is so much less than the unreliability due to the small number of cases as to be negligible. If the shape of a distribution is rectangular or approximately so, rank correlation is generally the better measure of the relationship existing.

## Assigning ranks

The first step in computing a coefficient of rank correlation is to assign ranks to the measures in the two series. To illus-

trate how this is done, a set of twenty-five scores, given in Table XXV, will be used. They are arranged in order of size, a convenient but not necessary procedure. Insofar as the calculation of rank correlation is concerned, it is immaterial whether rank 1 is given to the highest or to the lowest score,

TABLE XXV
ASSIGNMENT OF RANKS TO SCORES

| Score | Rank | Score | Rank | Score | Rank |
|---|---|---|---|---|---|
| 97 | 25 | 86 | 16 | 78 | 8.5 |
| 95 | 24 | 86 | 16 | 74 | 7 |
| 94 | 22.5 | 86 | 16 | 73 | 6 |
| 94 | 22.5 | 85 | 14 | 71 | 5 |
| 91 | 21 | 83 | 13 | 70 | 4 |
| 90 | 20 | 82 | 12 | 65 | 3 |
| 89 | 19 | 81 | 11 | 62 | 2 |
| 88 | 18 | 79 | 10 | 57 | 1 |
| | | 78 | 8.5 | | |

but it is consistent with percentile and other *-ile* ranks to give it to the lowest, so that practice is followed here.

The ranks assigned to the given scores are shown in the second of each pair of columns in the table. The lowest score, 57, receives rank 1; the next, 62, rank 2; and so on up to 74, which receives rank 7. The next two scores are both 78. In cases of such tied or equal scores the average of the ranks that would be assigned them if they were not tied is given to each. Here their ranks, if unequal, would be 8 and 9; hence 8.5, the average of 8 and 9, is given each. The next score, 79, receives 10, and so on until 85 is rank 14. The next three are tied at 86; hence each receives the average of 15, 16, and 17, which is 16. Another tie, of two cases, occurs at 94; hence each is ranked 22.5. Finally the highest is given rank 25, in other words, rank $N$.

## Computing coefficients of rank correlation

The easier but less reliable and in general less valid of the two methods is often called the *foot-rule method.* Its formula is

$$R = 1 - \frac{6\Sigma g}{N^2 - 1},$$

in which $R$ is the symbol for the coefficient of rank correlation and $g$ for the gains or positive differences in rank from one series to the other. The better but more difficult method employs the formula

$$\rho = 1 - \frac{6\Sigma D^2}{N(N^2 - 1)},$$

in which $\rho$ (rho) is the coefficient of rank correlation and $D$ the difference in rank, whether positive or negative. The values obtained for $R$ range from $+1.00$ down to $-.50$ if the number of cases is odd and to slightly more than that, depending on the number of cases, if it is even. Those for $\rho$ range from $+1.00$ down to $-1.00$ and generally are close to those of product-moment $r$ for the same data. As will be explained more fully in the next section, both $R$ and $\rho$ may be transmuted into approximately equivalent values of $r$.

Table XXVI presents the computation of both rank coefficients for the same data already employed for column product-moment correlation and elsewhere. The first two columns contain the scores themselves and the next two, headed $R_x$ and $R_y$, their ranks assigned according to the procedure explained in the last paragraph. The column headed $D$ contains the differences in rank, found by subtracting each entry in the $R_y$ column from the corresponding one in the $R_x$ column. The sum of the positive differences, or gains, is $\Sigma g$, here 28. It is advisable, although not necessary, to find the sum of the negative differences, or losses, also, as a check. Here $\Sigma l$ is $-28$, numerically equal to $\Sigma g$, as it always should be. Finally 28 is substituted in the formula, giving

$$R = 1 - \frac{6 \times 28.}{20^2 - 1} = + .579.$$

The last column of the table contains the squares of the entries in the $D$ column. Since all become positive through squaring, they are totalled into a single sum, 249. Using this in the second of the two formulas given above, we have

$$\rho = 1 - \frac{6 \times 249.}{20(20^2 - 1)} = + .813.$$

TABLE XXVI

COMPUTATION OF COEFFICIENTS OF RANK CORRELATION

| $X$ | $Y$ | $R_x$ | $R_y$ | $D$ | $D^2$ |
|---|---|---|---|---|---|
| 135 | 92 | 20 | 18 | +2. | 4. |
| 131 | 97 | 19 | 20 | −1. | 1. |
| 126 | 95 | 18 | 19 | −1. | 1. |
| 119 | 77 | 17 | 9 | +8. | 64. |
| 115 | 84 | 16 | 15 | +1. | 1. |
| 114 | 89 | 15 | 17 | −2. | 4. |
| 108 | 85 | 14 | 16 | −2. | 4. |
| 107 | 78 | 13 | 10 | +3. | 9. |
| 104 | 83 | 11.5 | 14 | −2.5 | 6.25 |
| 104 | 71 | 11.5 | 4 | +7.5 | 56.25 |
| 101 | 81 | 10 | 13 | −3. | 9. |
| 99 | 76 | 9 | 8 | +1. | 1. |
| 98 | 80 | 8 | 12 | −4. | 16. |
| 96 | 72 | 7 | 5 | +2. | 4. |
| 92 | 75 | 5.5 | 7 | −1.5 | 2.25 |
| 92 | 70 | 5.5 | 3 | +2.5 | 6.25 |
| 89 | 79 | 4 | 11 | −7. | 49. |
| 88 | 73 | 3 | 6 | −3. | 9. |
| 86 | 55 | 2 | 1 | +1. | 1. |
| 81 | 64 | 1 | 2 | −1. | 1. |
| | | | | $+28.0 = \Sigma g$ | $249.00 = \Sigma D^2$ |
| | | | | $-28.0 = \Sigma l$ | |

$$R = 1 - \frac{6 \times 28.}{20^2 - 1} = +.579 \text{ or } 1 - .01504 \times 28. = +.579 \text{ (equivalent } r = +.81)$$

$$\rho = 1 - \frac{6 \times 249.}{20(20^2 - 1)} = +.813 \text{ or } 1 - .000752 \times 249. = +.813 \text{ (equivalent } r = +.83)$$

The computation of both $R$ and $\rho$ may be facilitated by the use of a table containing values of $\frac{6}{N^2 - 1}$ and $\frac{6}{N(N^2 - 1)}$ for various values of $N$. With such values at hand the worker merely multiplies the correct one by $\Sigma g$, if $R$ is to be found, and by $\Sigma D^2$, if $\rho$ is wanted, and subtracts the product from 1. Table XXVII presents these values for $N$ ranging from 10 to 40, inclusive. Thus, as shown at the bottom of Table XXVI, $R$ for the data therein may also be obtained by finding the entry

TABLE XXVII

VALUES OF $\frac{6}{N^2 - 1}$ AND $\frac{6}{N(N^2 - 1)}$ CORRESPONDING TO VALUES OF $N$ FROM 10 TO 40

| $N$ | $\frac{6}{N^2-1}$ | $\frac{6}{N(N^2-1)}$ | $N$ | $\frac{6}{N^2-1}$ | $\frac{6}{N(N^2-1)}$ | $N$ | $\frac{6}{N^2-1}$ | $\frac{6}{N(N^2-1)}$ |
|---|---|---|---|---|---|---|---|---|
| 10 | .06061 | .006061 | 20 | .01504 | .000752 | 30 | .00667 | .000222 |
| 11 | .05000 | .004545 | 21 | .01364 | .000649 | 31 | .00625 | .000202 |
| 12 | .04196 | .003497 | 22 | .01242 | .000565 | 32 | .00587 | .000183 |
| 13 | .03571 | .002747 | 23 | .01136 | .000494 | 33 | .00551 | .000167 |
| 14 | .03077 | .002198 | 24 | .01043 | .000435 | 34 | .00519 | .000153 |
| 15 | .02679 | .001786 | 25 | .00962 | .000385 | 35 | .00490 | .000140 |
| 16 | .02353 | .001471 | 26 | .00889 | .000342 | 36 | .00463 | .000129 |
| 17 | .02083 | .001225 | 27 | .00824 | .000305 | 37 | .00439 | .000119 |
| 18 | .01858 | .001032 | 28 | .00766 | .000274 | 38 | .00416 | .000109 |
| 19 | .01667 | .000877 | 29 | .00714 | .000246 | 39 | .00395 | .000101 |
| | | | | | | 40 | .00375 | .000094 |

for $\frac{6}{N^2 - 1}$ when $N = 20$ in Table XXVII and proceding as just stated. Since this entry is .01504,

$$R = 1 - .01504 \times 28. = +.579.$$

Similarly,

$$\rho = 1 - .000752 \times 249. = +.813.$$

## Changing $R$ and $\rho$ into $r$

As was mentioned earlier, means are available for transmuting $R$ and $\rho$ into $r$. The results are not usually exactly the same as the directly computed value of $r$ for the same data, but approximate it. The discrepancy is due to the fact that the transmutation formulas are based upon an assumption as to the shape of the distribution of scores that is rarely fulfilled. For the two series of ranks $\rho$ is just the same as $r$, but since the exact data rarely have the same rectangular distribution as do their ranks, $\rho$ for the ranks and $r$ for the exact data differ. The formulas used are trigonometric, being

$$r = 2 \cos \frac{\pi}{3}(1 - R) - 1$$

and

$$r = 2 \sin \frac{\pi}{6} \rho.$$

Since the use of these formulas is somewhat laborious, the values of $R$ and $\rho$ that correspond to values of $r$ at intervals of .01 are presented in Table XXVIII. The first column con-

TABLE XXVIII

VALUES OF $r$ CORRESPONDING TO VALUES OF $R$ AND $\rho$

| $r$* | Positive $R$ | Negative $R$ | Positive or Negative $\rho$ | $r$ | Positive $R$ | Negative $R$ | Positive or Negative $\rho$ | $r$ | Positive $R$ | Negative $R$ | Positive or Negative $\rho$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| .00 | .000 | .000 | .000 | .34 | .201 | .179 | .326 | .68 | .453 | .347 | .663 |
| .01 | .006 | .006 | .010 | .35 | .208 | .184 | .336 | .69 | .461 | .351 | .673 |
| .02 | .011 | .011 | .019 | .36 | .214 | .189 | .346 | .70 | .470 | .356 | .683 |
| .03 | .017 | .017 | .029 | .37 | .221 | .194 | .355 | .71 | .480 | .361 | .693 |
| .04 | .022 | .022 | .038 | .38 | .227 | .199 | .365 | .72 | .489 | .366 | .703 |
| .05 | .028 | .027 | .048 | .39 | .234 | .204 | .375 | .73 | .498 | .371 | .714 |
| .06 | .034 | .033 | .057 | .40 | .240 | .209 | .385 | .74 | .508 | .375 | .724 |
| .07 | .039 | .038 | .067 | .41 | .247 | .214 | .394 | .75 | .517 | .380 | .734 |
| .08 | .045 | .044 | .076 | .42 | .254 | .219 | .404 | .76 | .527 | .385 | .744 |
| .09 | .050 | .049 | .086 | .43 | .261 | .224 | .414 | .77 | .538 | .390 | .755 |
| .10 | .056 | .055 | .096 | .44 | .268 | .229 | .424 | .78 | .548 | .395 | .765 |
| .11 | .062 | .060 | .105 | .45 | .274 | .234 | .433 | .79 | .559 | .400 | .776 |
| .12 | .068 | .065 | .115 | .46 | .281 | .239 | .443 | .80 | .569 | .404 | .786 |
| .13 | .074 | .070 | .124 | .47 | .288 | .244 | .453 | .81 | .580 | .409 | .796 |
| .14 | .079 | .076 | .134 | .48 | .296 | .249 | .463 | .82 | .592 | .414 | .807 |
| .15 | .085 | .081 | .143 | .49 | .303 | .254 | .473 | .83 | .603 | .419 | .817 |
| .16 | .091 | .086 | .153 | .50 | .310 | .259 | .483 | .84 | .615 | .423 | .828 |
| .17 | .097 | .091 | .163 | .51 | .317 | .264 | .492 | .85 | .628 | .428 | .838 |
| .18 | .103 | .096 | .172 | .52 | .324 | .269 | .502 | .86 | .641 | .433 | .849 |
| .19 | .109 | .102 | .182 | .53 | .332 | .274 | .512 | .87 | .654 | .438 | .859 |
| .20 | .115 | .107 | .191 | .54 | .339 | .279 | .522 | .88 | .667 | .443 | .870 |
| .21 | .121 | .112 | .201 | .55 | .347 | .283 | .532 | .89 | .682 | .448 | .881 |
| .22 | .127 | .117 | .211 | .56 | .354 | .288 | .542 | .90 | .697 | .452 | .892 |
| .23 | .133 | .123 | .220 | .57 | .362 | .293 | .552 | .91 | .713 | .457 | .902 |
| .24 | .139 | .128 | .230 | .58 | .370 | .298 | .562 | .92 | .729 | .462 | .913 |
| .25 | .145 | .133 | .239 | .59 | .378 | .303 | .572 | .93 | .747 | .467 | .924 |
| .26 | .151 | .138 | .249 | .60 | .385 | .308 | .582 | .94 | .766 | .471 | .934 |
| .27 | .157 | .143 | .259 | .61 | .393 | .313 | .592 | .95 | .786 | .476 | .945 |
| .28 | .163 | .148 | .268 | .62 | .402 | .318 | .602 | .96 | .809 | .481 | .956 |
| .29 | .169 | .154 | .278 | .63 | .410 | .322 | .612 | .97 | .834 | .486 | .967 |
| .30 | .176 | .159 | .288 | .64 | .418 | .327 | .622 | .98 | .865 | .490 | .978 |
| .31 | .182 | .164 | .297 | .65 | .427 | .332 | .632 | .99 | .905 | .495 | .989 |
| .32 | .188 | .169 | .307 | .66 | .435 | .337 | .642 | 1.00 | 1.000 | .500 or more | 1.000 |
| .33 | .195 | .174 | .317 | .67 | .444 | .342 | .652 | | | | |

* The sign of $r$ is determined by that of $R$ or $\rho$.

tains the values of $r$; the second has the corresponding ones of $R$ when it is positive; the third has those of negative $R$, in which case the minus sign should be prefixed to the corresponding value of $r$; and the fourth those of $\rho$, for which its sign, plus or minus as the case may be, should be prefixed to $r$.

From Table XXVIII it appears that for $R = +.579$, the value found in Table XXVI, the corresponding value of $r$ is $+.81$, and that for $\rho = +.813$, $r = +.83$. These are both fairly close to $+.84$, the value of $r$ found directly by the product-moment method.

## EXERCISES AND PROBLEMS

**1.** Compute both coefficients of rank correlation for each set of data in Exercise 1 on page 134.

**2.** Find the value of $r$ corresponding to each of the rank coefficients computed in Exercise 1.

## REFERENCES

Garrett, Henry E., *Statistics in Psychology and Education.* Second Edition. Pages 359–366. New York: Longmans, Green & Company, Inc., 1937.

Guilford, J. P., *Fundamental Statistics in Psychology and Education.* Pages 227–231. New York: McGraw-Hill Book Company, Inc., 1942.

Kenney, John F., *Mathematics of Statistics.* Part I, pages 209–212. New York: D. Van Nostrand Company, Inc., 1939.

Lindquist, E. F., *Statistical Analysis in Educational Research.* Pages 246–248. Boston: Houghton Mifflin Company, 1940.

Mode, Elmer B., *The Elements of Statistics.* Pages 307–312. New York: Prentice-Hall, Inc., 1941.

Peters, Charles C., and Van Voorhis, Walter R., *Statistical Procedures and Their Mathematical Bases.* Pages 103–109. New York: McGraw-Hill Book Company, Inc., 1940.

Walker, Helen M., *Elementary Statistical Methods.* Pages 74–78, 242–245. New York: Henry Holt & Company, Inc., 1943.

# CHAPTER XI

# The Correlation and Regression of More Than Two Variables

## Partial correlation

Partial and multiple correlation and regression, that is, correlation and regression involving more than two variables, is too large a topic to be treated at all adequately in an elementary text. It seems desirable, however, that the chief functions of such procedures be stated and their natures explained briefly.

The *coefficient of partial,* or *net, correlation* is an estimate of what the ordinary coefficient of correlation between two variables would be if the portion of the total correlation associated with one or more other variables were eliminated. In other words, it is an estimate of what the coefficient would be if the values of the one or more others were held constant. For example, the correlation between height and weight of a group of children in Grades I–VIII is associated with the fact that both height and weight correlate with age; hence, if this portion of the total correlation is removed, that remaining is less. The result so obtained is an estimate of the average coefficient that would be found between height and weight for a number of subgroups, each of which contained children all of the same age.

Although the technique of partial correlation is frequently valuable in investigations and studies concerned with cause, the evidence it offers never proves causal relationship, but only association. There are also other limitations connected with its use, largely because we usually do not know enough about the nature of the intercorrelations among the variables concerned; so much caution should be exercised in its interpreta-

tion. It is only when the variable or variables eliminated act as a single factor or unified group of factors upon the two between which the correlation is desired that the partial coefficient is really a valid estimate of what the ordinary coefficient would be if the one or more others were held constant. Moreover, since it is based upon product-moment coefficients, it measures only rectilinear relationship.

The symbol for a coefficient of partial correlation is $r$ with a subscript of as many symbols as there are variables involved. The first two symbols denote the two variables between which the partial correlation is found, and the one or more following them denote the variable or variables eliminated. A point halfway up is used to separate the first two from the other or others. Thus $r_{12\cdot3}$ means the coefficient of partial correlation between variables 1 and 2 when variable 3 is eliminated; $r_{13\cdot24}$ that between variables 1 and 3 when 2 and 4 are eliminated; and so on. Coefficients are referred to as being of the order represented by the number of symbols after the dot. Thus ordinary coefficients, which have none, are *zero-order coefficients;* those such as $r_{12\cdot3}$ and $r_{xz\cdot y}$ are *first order;* those such as $r_{12\cdot34}$ and $r_{AC\cdot BD}$ are *second order;* and so on.

The general formula for finding any partial coefficient is in terms of coefficients of the next lower order. For any number, $n$, of variables, it is

$$r_{12\cdot34\ldots n} = \frac{r_{12\cdot34\ldots(n-1)} - r_{1n\cdot34\ldots(n-1)}r_{2n\cdot34\ldots(n-1)}}{\sqrt{(1 - r^2_{1n\cdot34\ldots(n-1)})(1 - r^2_{2n\cdot34\ldots(n-1)})}}.$$

The denominator may also be written

$$k_{1n\cdot34\ldots(n-1)}k_{2n\cdot34\ldots(n-1)}.$$

For only three variables it becomes

$$r_{12\cdot3} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{(1 - r_{13}^2)(1 - r_{23}^2)}}$$

or

$$\frac{r_{12} - r_{13}r_{23}}{k_{13}k_{23}}.$$

Since the order of the subscripts after the dot, as well as that of those before the dot, is not significant, the formula for

second-order coefficients may be written in either one of two ways,

$$r_{12\cdot34} \text{ or } r_{12\cdot43} = \frac{r_{12\cdot3} - r_{14\cdot3}r_{24\cdot3}}{k_{14\cdot3}k_{24\cdot3}}$$

or

$$\frac{r_{12\cdot4} - r_{13\cdot4}r_{23\cdot4}}{k_{13\cdot4}k_{23\cdot4}},$$

that for third-order coefficients in any one of three, and so on. In such cases the use of at least two ways provides a check upon the result. Anyone planning to compute many partial coefficients should consult some more extensive treatment of the topic for suggestions as to convenient and economical organization of the work.

To illustrate the use of the formula, we may assume that a coefficient of $+.82$ has been found between height and weight, one of $+.85$ between height and age, one of $+.76$ between weight and age, and that we wish the partial coefficient between height and weight with age eliminated. Using the initial letters as subscripts, we have

$$r_{HW\cdot A} = \frac{r_{HW} - r_{HA}r_{WA}}{\sqrt{(1 - r_{HA}^2)(1 - r_{WA}^2)}} = \frac{.82 - .85 \times .76}{\sqrt{(1 - .85^2)(1 - .76^2)}} = + .51.$$

In other words, .51 is the estimated correlation between height and weight when age is constant. It is also possible to find the partial coefficient of height with age when weight is eliminated and of weight with age when height is eliminated. Thus

$$r_{HA\cdot W} = \frac{.85 - .82 \times .76}{\sqrt{(1 - .82^2)(1 - .76^2)}} = + .61$$

and

$$r_{WA\cdot H} = \frac{.76 - .85 \times .82}{\sqrt{(1 - .85^2)(1 - .82^2)}} = + .21.$$

In most cases partial coefficients are smaller than the corresponding zero-order coefficients, but they may be larger, especially when one of the coefficients is negative. Thus if $r_{12} = +.60$, $r_{13} = +.20$, and $r_{23} = +.80$,

$$r_{12\cdot3} = \frac{.60 - .20 \times .80}{\sqrt{(1 - .20^2)(1 - .80^2)}} = + .75,$$

and if $r_{12} = +.60$, $r_{13} = -.20$, and $r_{23} = +.40$,

$$r_{12\cdot3} = \frac{.60 - (-.20 \times .40)}{\sqrt{(1 - .20^2)(1 - .40^2)}} = +.76,$$

both larger than $+.60$. Their range is the same as that of zero-order coefficients, from $-1.00$ to $+1.00$.

The reader should not confuse partial correlation with part correlation, semipartial correlation, and other varieties which have some characteristics in common with it, but differ in others. They are much less frequently employed.

## Multiple correlation

The *coefficient of multiple correlation* is the coefficient of correlation between one variable and two or more others combined or weighted so as to yield the maximum relationship possible from an additive, as contrasted with a multiplicative, exponential, or other, combination of them. It is based upon zero-order coefficients, and hence measures rectilinear relationship only. It is always positive, ranging from 0 up to $+1.00$, and therefore indicates the amount of relationship without regard to direction.

The formula for a coefficient of multiple correlation can be written in terms of zero-order coefficients, but in this form becomes very unwieldy for more than three or four variables. A simpler form is in terms of partial correlation or partial alienation coefficients. In situations involving more than a few variables, computation by these methods becomes very laborious. A number of improved methods are available. Of these, the one commonly called the Doolittle Method is probably the best for most workers. Explanation of it, as well as other helpful suggestions, may be found in some of the references at the end of this chapter.

The general formula for multiple correlation may be written

$R_{1\cdot23\ldots n}$ or $R_{1(23\ldots n)} =$

$$\sqrt{1 - (1 - r_{12}^2)(1 - r_{13\cdot2}^2)(1 - r_{14\cdot23}^2)\ldots\ldots(1 - r_{1n\cdot23\ldots(n-1)}^2)}$$

or

$$\sqrt{1 - k_{12}^2 k_{13\cdot2}^2 k_{14\cdot23}^2 \ldots\ldots k_{1n\cdot23\ldots(n-1)}^2}.$$

When only three variables are concerned, that is, when one is correlated with the combination of two others, it reduces to

$$R_{1\cdot 23} = \sqrt{1 - (1 - r_{12}^2)(1 - r_{13\cdot 2}^2)}$$

or

$$\sqrt{1 - k_{12}^2 k_{13\cdot 2}^2},$$

or, in terms of zero-order coefficients,

$$\sqrt{\frac{r_{12}^2 + r_{13}^2 - 2r_{12}r_{13}r_{23}}{1 - r_{23}^2}}.$$

Since the order of subscripts after the point is immaterial—so that $R_{1\cdot 23}$ and $R_{1\cdot 32}$ are the same, as likewise are $R_{1\cdot 234}$, $R_{1\cdot 243}$, $R_{1\cdot 324}$, $R_{1\cdot 342}$, $R_{1\cdot 423}$, and $R_{1\cdot 432}$, and so on in increasing number for more variables—multiple coefficients may be computed in two or more ways in order to check. Thus for $R_{1\cdot 23}$ or $R_{1\cdot 32}$, which are the same, we may use

$$\sqrt{1 - (1 - r_{13}^2)(1 - r_{12\cdot 3}^2)}$$

instead of the form given above.

To illustrate the computation of a multiple coefficient for one variable with two others, the data employed in the preceding section may be used. It will be recalled that the given zero-order coefficients were $r_{HW} = +.82$, $r_{HA} = +.85$, and $r_{WA} = +.76$, from which it was found that $r_{HW\cdot A} = +.51$, $r_{HA\cdot W} = +.61$ and $r_{WA\cdot H} = +.21$. From these, we have

$$R_{H\cdot WA} = \sqrt{1 - (1 - .82^2)(1 - .61^2)}$$

or

$$\sqrt{\frac{.82^2 + .85^2 - 2 \times .82 \times .85 \times .76}{1 - .76^2}}$$

or, using it as $R_{H\cdot AW}$,

$$\sqrt{1 - (1 - .85^2)(1 - .51^2)},$$

each of which equals .89. Similarly,

$$R_{W\cdot HA} = \sqrt{1 - (1 - .82^2)(1 - .21^2)}$$

or

$$\sqrt{1 - (1 - .76^2)(1 - .51^2)}$$

or

$$\sqrt{\frac{.82^2 + .76^2 - 2 \times .82 \times .76 \times .85}{1 - .85^2}} = .83;$$

and

$$R_{A \cdot HW} = \sqrt{1 - (1 - .85^2)(1 - .21^2)}$$

or

$$\sqrt{1 - (1 - .76^2)(1 - .61^2)}$$

or.

$$\sqrt{\frac{.85^2 + .76^2 - 2 \times .85 \times .76 \times .82}{1 - .82^2}} = .86.$$

Each of the multiple coefficients just found is only slightly greater than the highest zero-order coefficient of the dependent variable with one of the independent ones. In other words, the total degree of relationship is increased only slightly by adding the influence of the second independent variable. The reason for this is that the two independent variables, in each case, are rather closely correlated with each other. For the same coefficients with the dependent variable, the lower the correlation among the independent variables, the higher the multiple coefficient. Thus, for example, if $r_{12} = +.60$, $r_{13} = +.60$, and $r_{23} = +.80$, $R_{1 \cdot 23} = .63$; but if $r_{23}$ drops to $+.60$, $R_{1 \cdot 23} = .67$; if $r_{23} = +.40$, $R_{1 \cdot 23} = .72$; if $r_{23} = +.20$, $R_{1 \cdot 23} = .78$; and if $r_{23} = .00$, $R_{1 \cdot 23} = .85$. Multiple $R$ can never be less than zero-order $r$.

Quite often a multiple coefficient found from one set of data is considered to apply to one or more comparable sets of data or to the total population, of which a sample was used to secure the obtained coefficient. In both cases the value obtained is larger than the true value for the other situation. Generally the difference is so small—perhaps .01 or .02—that it is neglected; but, if many variables are concerned, the number of cases rather small, and the multiple coefficient low, the difference—or *shrinkage*, as it is called—may be large enough to be significant. Quite similar, but not identical, correction formula based upon the number of cases and of variables are available

for the two types of situations. That for estimating the multiple coefficient for a comparable situation is

$$R_{corr} = \sqrt{1 - \frac{1 - R^2}{1 - \frac{n}{N}}} \quad \text{or} \quad \sqrt{\frac{NR^2 - n}{N - n}}$$

and that for estimating the multiple coefficient for the total population is

$$R_{corr} = \sqrt{1 - (1 - R^2)\frac{N - 1}{N - n}} \quad \text{or} \quad \sqrt{\frac{1 - n + R^2(N - 1)}{N - n}},$$

in which $R$ is the obtained multiple coefficient, $N$ the number of cases, and $n$ the number of variables. To illustrate their application, let us take $R = .60$ for 50 cases and 5 variables. For a comparable situation,

$$R_{corr} = \sqrt{\frac{50 \times .60^2 - 5}{50 - 5}} = .54$$

and, for a total population of which the 50 cases constitute a sample,

$$R_{corr} = \sqrt{\frac{1 - 5 + .60^2(50 - 1)}{50 - 5}} = .55.$$

It is evident that the correction formula is worth applying in such cases. If, however, the obtained coefficient is .80, the number of cases 100, and that of variables only 4, the formulas give, respectively,

$$R_{corr} = \sqrt{\frac{100 \times .80^2 - 4}{100 - 4}} = .79$$

and

$$R_{corr} = \sqrt{\frac{1 - 4 + .80^2(100 - 1)}{100 - 4}} = .79,$$

both only .01 less than the obtained value of $R$.

## Partial and multiple regression

Partial and multiple regression bear the same relationship to partial and multiple correlation as does ordinary regression to zero-order correlation. *Partial* or *net regression* refers to the best rectilinear means of estimating a dependent variable from

one of two or more independent ones when the relationship associated with the other independent one or ones is eliminated. *Multiple regression* refers to the best rectilinear means of estimating a dependent variable from a combination of two or more independent ones. Whether the equation employed is considered as being a partial regression equation or a multiple regression equation depends upon the desired emphasis—whether it is on the separate terms thereof or on the whole equation. In effect, partial regression equation and multiple regression equation are synonymous. Such an equation should be employed whenever we wish to estimate one variable from two or more others with which it is correlated and are content to use a rectilinear relationship.

Such equations are similar to those of zero-order regression except for having as many terms as there are variables concerned and for employing partial correlation coefficients and standard deviations instead of ordinary ones. Since they use the latter, methods of computing them will be given.

A *partial standard deviation* is simply the standard deviation of the variable involved when the effects of others included in partial correlation are eliminated. Its formula may be given in any of several forms, of which the following are the most frequently employed:

$$\sigma_{1\cdot234\cdots n} = \sigma_1\sqrt{(1 - r_{12}^2)(1 - r_{13\cdot2}^2)(1 - r_{14\cdot23}^2)\cdots\cdots(1 - r_{1n\cdot23\cdots(n-1)}^2)},$$

$$\sigma_1 k_{12} k_{13\cdot2} k_{14\cdot23}\cdots\cdots k_{1n\cdot23\cdots(n-1)},$$

and

$$\sigma_1\sqrt{1 - R_{1\cdot234\cdots n}^2}\,.$$

For two variables the formula becomes

$$\sigma_{1\cdot2} = \sigma_1\sqrt{1 - r_{12}^2} \text{ or } \sigma_1 k_{12}$$

and for three it is

$$\sigma_{1\cdot23} = \sigma_1\sqrt{(1 - r_{12}^2)(1 - r_{13\cdot2}^2)},$$

$$\sigma_1 k_{12} k_{13\cdot2},$$

$$\sigma_{1\cdot2}\sqrt{1 - r_{13\cdot2}^2},$$

or

$$\sigma_1\sqrt{1 - R_{1\cdot23}^2}.$$

For three or more variables there are increasing numbers of variations which may be used, just as with partial coefficients. Thus, since $\sigma_{1\cdot 23}$ and $\sigma_{1\cdot 32}$ are the same, the formula for them may be

$$\sigma_1\sqrt{(1 - r_{13}^2)(1 - r_{12.3}^2)},$$

$$\sigma_1 k_{13} k_{12\cdot 3},$$

or

$$\sigma_{1\cdot 3}\sqrt{1 - r_{12\cdot 3}^2}$$

as well as the forms already given.

The partial regression coefficients are found by a procedure analogous to that for zero-order coefficients; hence

$$b_{12\cdot 34\cdots n} = r_{12\cdot 34\cdots n}\,\frac{\sigma_{1\cdot 34\cdots n}}{\sigma_{2\cdot 34\cdots n}}.$$

For three variables this becomes

$$b_{12\cdot 3} = r_{12\cdot 3}\left(\frac{\sigma_{1\cdot 3}}{\sigma_{2\cdot 3}}\right),$$

which may be changed to

$$r_{12\cdot 3}\left(\frac{\sigma_1\sqrt{1 - r_{13}^2}}{\sigma_2\sqrt{1 - r_{23}^2}}\right) \quad \text{or} \quad \frac{(r_{12} - r_{13}\cdot r_{23})\sigma_1}{(1 - r_{23}^2)\sigma_2},$$

forms sometimes more convenient to use.

After partial regression coefficients have been found, they may be employed to make partial or multiple regression equations by the same procedure followed for ordinary regression equations, the only difference being the presence of more than one independent variable. Thus in general, if $X_1$, $X_2$, $X_3$, and so on to $X_n$ are the variables involved, the equation is

$$X_1 = b_{12\cdot 34\cdots n}(X_2 - M_2) + b_{13\cdot 24\cdots n}(X_3 - M_3) + \\ \cdots\cdot b_{1n\cdot 23\cdots\cdot(n-1)}(X_n - M_n) + M_1.$$

The use of such an equation may be illustrated by data secured for predicting college marks from high-school marks and other data. The data for three variables—high-school per-cent mark, intelligence test score, and college per-cent mark—were as follows, using the initial letters to designate them: $M_H = 84.$,

$\sigma_H = 8.5, M_I = 54., \sigma_I = 10.2, M_C = 82., \sigma_C = 9.6, r_{HI} = +.64,$ $r_{CH} = +.55$, $r_{CI} = +.52$. The equation desired is that for estimating college marks from high-school marks and intelligence test scores, that is,

$$C = b_{CH \cdot I}(H - \mathrm{M}_H) + b_{CI \cdot H}(I - M_I) + M_C.$$

Since

$$b_{CH \cdot I} = \frac{(r_{CH} - r_{CI}r_{HI})\sigma_C}{(1 - r^2_{HI})\sigma_H} = \frac{(.55 - .52 \times .64)9.6}{(1 - .64^2)8.5} = .4155$$

and

$$b_{CI \cdot H} = \frac{(r_{CI} - r_{CH}r_{IH})\sigma_C}{(1 - r^2_{IH})\sigma_I} = \frac{(.52 - .55 \times .64)9.6}{(1 - .64^2)10.2} = .2678,$$

the equation is

$$C = .4155(H - 84.) + .2678(I - 54.) + 82.,$$

which reduces to

$$C = .42H + .27I + 32.6.$$

Stated verbally, this means that the best prediction of a college mark is obtained by adding .42 of high-school mark, .27 of intelligence test score, and 32.6.

In case there is no need to estimate actual scores, but only relative ones—a situation that sometimes exists—an easier formula and method may be employed. If the dependent variable is denoted by *1*, the independent ones by *2* and *3*, the weight to be given variable *3* when *2* is weighted 1.00 is given by

$$W = \frac{\sigma_2(r_{13} - r_{12}r_{23})}{\sigma_3(r_{12} - r_{13}r_{23})}.$$

In this case, calling college mark *1*, high-school mark *2*, and intelligence test score *3*,

$$W = \frac{\sigma_H(r_{CI} - r_{CH}r_{HI})}{\sigma_I(r_{CH} - r_{CI}r_{HI})} = \frac{8.5(.52 - .55 \times .64)}{10.2(.55 - .52 \times .64)} = .64.$$

Hence, if high-school mark is weighted 1.00, intelligence score is weighted .64. This is, as it must be, the ratio between the corresponding weights, .27 and .42, yielded by the regression equation.

The interpretation of partial and multiple correlation and regression is similar to that of zero-order correlation and regression. Partial $k$ has already been introduced, and multiple $K$ and partial and multiple $E_m$, $g$, and $E_g$ may be similarly interpreted. Errors of estimate may also be found in the same way. When obtained, all these measures of relationship and error are subject to interpretations similar to those given in Chapter VIII.

## EXERCISES AND PROBLEMS

**1.** Compute the coefficients of partial and multiple correlation, and the partial or multiple regression equations, for each of the following sets of data:

(*a*)

| Algebra Mark | Arithmetic Mark | Prognostic Test Score |
|---|---|---|
| 83 | 92 | 71 |
| 79 | 83 | 62 |
| 91 | 96 | 76 |
| 85 | 83 | 64 |
| 88 | 93 | 66 |
| 72 | 76 | 53 |
| 59 | 65 | 42 |
| 86 | 88 | 70 |
| 77 | 83 | 61 |
| 75 | 75 | 62 |
| 95 | 94 | 81 |
| 82 | 87 | 77 |
| 68 | 73 | 48 |
| 74 | 82 | 60 |
| 87 | 93 | 74 |
| 71 | 72 | 57 |
| 85 | 86 | 69 |
| 90 | 89 | 73 |
| 78 | 84 | 63 |
| 84 | 86 | 65 |

(*b*)

$M_1 = 48$, $M_2 = 29$, $M_3 = 74$,
$\sigma_1 = 7.2$, $\sigma_2 = 5.5$, $\sigma_3 = 8.8$,
$r_{12} = +.48$, $r_{13} = +.75$, $r_{23} = +.33$.

(*c*)

$M_1 = 35$, $M_2 = 76$, $M_3 = 57$, $M_4 = 28$,
$\sigma_1 = 4.2$, $\sigma_2 = 12.6$, $\sigma_3 = 6.6$, $\sigma_4 = 3.9$,
$r_{12} = +.65$, $r_{13} = +.58$, $r_{14} = +.49$,
$r_{23} = +.36$, $r_{24} = +.44$, $r_{34} = .22$.

**2.** Compute $W$, as used in the last section, for Parts (*a*) and (*b*) of Exercise 1.

## REFERENCES

Ezekiel, Mordecai, *Methods of Correlation Analysis.* Second Edition. Chaps. X–XVI, XXI–XXII; pages 450–460, 474–485. New York: John Wiley & Sons, Inc., 1941.

Fisher, R. A., *Statistical Methods for Research Workers.* Eighth Edition. Pages 180–185, 252–254. Edinburgh: Oliver and Boyd, 1941.

Garrett, Henry E., *Statistics in Psychology and Education.* Second Edition. Chap. XIV. New York: Longmans, Green & Company, Inc., 1937.

Guilford, J. P., *Fundamental Statistics in Psychology and Education.* Chap. XIII. New York: McGraw-Hill Book Company, Inc., 1942.

Kenney, John P., *Mathematics of Statistics.* Part II, Chap. V. New York: D. Van Nostrand Company, Inc., 1939.

Lindquist, E. F., *Statistical Analysis in Educational Research.* Pages 248–257. Boston: Houghton Mifflin Company, 1940.

Mode, Elmer B., *The Elements of Statistics.* Pages 312–314. New York: Prentice-Hall, Inc., 1941.

Monroe, Walter S., and Engelhart, Max D., *The Scientific Study of Educational Problems.* Pages 377–382. New York: The Macmillan Company, 1926.

Peters, Charles C., and Van Voorhis, Walter R., *Statistical Procedures and Their Mathematical Bases.* Chap. VIII. New York: McGraw-Hill Book Company, Inc., 1940.

# CHAPTER XII

# Other Methods of Correlation

## Ratios of correlation

*Ratios of correlation*, sometimes known as *coefficients of curvilinear correlation*, are the most commonly employed measures of curvilinear correlation. They are found from double-entry tables such as are used for coefficients of correlation and make use of most of the same steps plus a few additional ones. For each such table with two variables there are two ratios, one for each variable on the other. In cases where a variable and an attribute are correlated, a ratio may be found for the variable on the attribute, but not vice versa.

A ratio of correlation summarizes the deviations of the data from the curved line that best fits the means of the rows or columns, according to which ratio it is, just as a coefficient summarizes the deviations from the straight line that best fits both sets of means. It is always positive, ranging from 0 to +1.00. It cannot be less than the coefficient of correlation for the same data.

The formula for a ratio, abbreviated $\eta$ (eta), may be given in various forms. When, as usually, the coefficient is also desired, the following is probably most convenient:

$$\eta_{xy} = \frac{\sqrt{\dfrac{\Sigma\dfrac{(\Sigma x)^2}{f_y}}{N} - c_x^2}}{\sigma_x} \text{ and } \eta_{yx} = \frac{\sqrt{\dfrac{\Sigma\dfrac{(\Sigma y)^2}{f_x}}{N} - c_y^2}}{\sigma_y}.$$

Another form is sometimes preferable when the coefficient is not wanted, but is so rarely employed that it will not be given here.

The computation of the ratios for the data employed in Table XXI is given in Table XXIX. This table repeats the

TABLE

COMPUTATION OF THE RATIOS OF

| Final Exam. Score | Total Score on | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | 120– | 140– | 160– | 180– | 200– | 220– | 240– | 260– | 280– |
| 130– | | | | | | | | | |
| 125– | | | | | | | | | |
| 120– | | | | | | | | | 1 |
| 115– | | | | | | 1 | 1 | 1 | 2 |
| 110– | | | | | | | 1 | 1 | |
| 105– | | | | 2 | | 1 | 2 | 3 | 1 |
| 100– | | | | | 2 | 2 | 2 | 2 | |
| 95– | | | 1 | 2 | 2 | 2 | | | |
| 90– | | | 2 | 1 | 1 | | | | |
| 85– | | | | | | | | | |
| 80– | | | | | | | | | |
| 75– | 2 | 1 | 1 | | | | | | |
| $f_x$ | 2 | 1 | 4 | 5 | 5 | 6 | 6 | 7 | 4 |
| $x$ | −6 | −5 | −4 | −3 | −2 | −1 | 0 | +1 | +2 |
| $fx$ | −12 | −5 | −16 | −15 | −10 | −6 | −64 | +7 | +8 |
| $fx^2$ | 72 | 25 | 64 | 45 | 20 | 6 | 0 | 7 | 16 |
| $\Sigma y$ | −12 | −6 | −14 | −7 | −9 | −4 | +1 | +1 | +7 |
| $\Sigma xy$ | +72 | +30 | +56 | +21 | +18 | +4 | 0 | +1 | +14 |
| $(\Sigma y)^2$ | 144 | 36 | 196 | 49 | 81 | 16 | 1 | 1 | 49 |
| $\frac{(\Sigma y)^2}{f_x}$ | 72.00 | 36.00 | 49.00 | 9.80 | 16.20 | 2.67 | .17 | .14 | 12.25 |

$$\eta_{xy} = \frac{\sqrt{\frac{341.46}{50} - .0324}}{2.889} = .902 \qquad \eta_{yx} = \frac{\sqrt{\frac{330.23}{50} - .0196}}{2.807} = .914$$

XXIX

Correlation Between Two Variables

| Short Tests | | | | $y$ | $fy$ | $fy^2$ | $\Sigma x$ | $\Sigma xy$ | $(\Sigma x)^2$ | $\frac{(\Sigma x)^2}{f_y}$ |
|---|---|---|---|---|---|---|---|---|---|---|
| 300– | 320– | 340– | $f_y$ | | | | | | | |
| | 2 | 1 | 3 | +5 | +15 | 75 | +13 | +65 | 169 | 56.33 |
| 1 | 2 | 1 | 4 | +4 | +16 | 64 | +16 | +64 | 256 | 64.00 |
| | | | 1 | +3 | + 3 | 9 | + 2 | + 6 | 4 | 4.00 |
| 1 | 1 | | 7 | +2 | +14 | 28 | +11 | +22 | 121 | 17.29 |
| | 1 | | 3 | +1 | + 3 | 3 | + 5 | + 5 | 25 | 8.33 |
| | | | 9 | 0 | +51 | 0 | − 2 | 0 | 4 | .44 |
| | | | 8 | −1 | − 8 | 8 | − 4 | + 4 | 16 | 2.00 |
| | | | 7 | −2 | −14 | 28 | −16 | +32 | 256 | 36.57 |
| | | | 4 | −3 | −12 | 36 | −13 | +39 | 169 | 42.25 |
| | | | 0 | −4 | 0 | 0 | 0 | 0 | 0 | .00 |
| | | | 0 | −5 | 0 | 0 | 0 | 0 | 0 | .00 |
| | | | 4 | −6 | −24 | 144 | −21 | +126 | 441 | 110.25 |
| 2 | 6 | 2 | 50 | | −58<br>− 7 | 395 | +47<br>−56<br>− 9 | +363 | | 341.46 |
| +3 | +4 | +5 | | | | | | | | |
| +6 | +24 | +10 | +55 | − 9 | | | | | | |
| 18 | 96 | 50 | 419 | | | | | | | |
| +6 | +21 | +9 | −52+45 | −7 | | | | | | |
| +18 | +84 | +45 | +363 | | | | | | | |
| 36 | 441 | 81 | | | | | | | | |
| 18.00 | 73.50 | 40.50 | 330.23 | | | | | | | |

$$c_x = \frac{-9}{50} = -.18 \qquad c_y = \frac{-7}{50} = -.14$$

$$S_x^2 = \frac{419}{50} = 8.3800 \qquad S_y^2 = \frac{395}{50} = 7.9000$$

$$c_x^2 = .0324 \qquad c_y^2 = .0196$$

$$\sigma_x^2 = 8.3476 \qquad \sigma_y^2 = 7.8804$$

$$\sigma_x = 2.889 \qquad \sigma_y = 2.807$$

$$\frac{363}{50} = 7.2600$$

$$c_x c_y = .0252$$

$$p = 7.2348$$

$$r = \frac{7.2348}{2.889 \times 2.807} = +.892$$

former one, except that one method of finding $r$ is omitted, and it also includes the two additional rows and columns needed to calculate the ratios and the substitution in the formulas therefor. The first added column, headed $(\Sigma x)^2$, contains the squares of the entries in the $\Sigma x$ column, and the next contains the quotients of these squares divided by the corresponding frequencies. Thus, since for the top row $\Sigma x$ is 13, the entry in $(\Sigma x)^2$ column is $13^2$ or 169 and that in the final column is 169 divided by 3, the frequency for that row, which gives 56.33. This last column is then totalled, the result being 341.46, and this along with $N$, $\sigma_x$, and $c_x^2$ substituted in the formula, giving

$$\eta_{xy} = \frac{\sqrt{\frac{341.46}{50} - .0324}}{2.889} = .902.$$

Similarly the entries in the two rows at the bottom of the table are found, the last one summed to yield 330.23, and $\eta_{yx}$ found to be

$$\frac{\sqrt{\frac{330.23}{50} - .0196}}{2.807} = .914.$$

Since these values are only slightly larger than that of the coefficient, $+.892$, it appears that the latter is almost as good a measure of the existing relationship as is either ratio. A number of suggestions for the interpretation of differences between ratios and coefficients have been advanced. The most widely known is Blakeman's *criterion of linearity* ($C$ *of* $L$). It is based upon the standard error of $\eta^2 - r^2$. It has been shown, however, that Blakeman's test may lead to erroneous conclusions. The use of the $\chi^2$ test for correctness of fit, which is explained in Chapter XIII, appears decidedly preferable. Its use for this purpose will be illustrated there. For still better understanding of the significance of ratios of correlation, readers should employ the tables prepared by Woo.[1]

[1] Woo, T. L., "Tables for Ascertaining the Significance or Non-significance of Association Measured by the Correlation Ratio." *Biometrika*, Vol. XXI, pages 1–66. December, 1929. Also in Pearson, Karl, Editor, *Tables for Statisticians and Biometricians*. Part II, pages xl–xliv, 16–72. Cambridge: Cambridge University Press, 1931.

The use of the ratio of correlation when one variable and one attribute are involved is shown in Table XXX, which contains data on the intelligence quotients of boys preferring certain types of books. In such cases the attribute may be grouped into any number of classes from two up. Here there are six. The order of the arrangement of the classes of the attribute is immaterial, since it does not affect the result. The computation is the same as that in Table XXIX, except that a part of what is there is not found here. Finding $c_y^2$ and $\sigma_y$, and $\frac{(\Sigma y)^2}{f_x}$ for each array, and summing the latter, then using them in the formula, we have

$$\eta_{yx} = \frac{\sqrt{\frac{40.23}{95} - .0004}}{1.61} = .40.$$

Ratios of correlation obtained as described above are subject to several errors, the chief of which is connected with the number of classes. If there are as many classes as cases, the ratio is always 1.00. In any instance, the more classes, the larger the error. An approximate correction for this error may be made by employing the formula

$$\eta_{corr} = \sqrt{\frac{\eta^2 - \frac{n-3}{N}}{1 - \frac{n-3}{N}}},$$

in which $n$ is the number of classes of the first term in the subscript of $\eta$. Applying this to the data in Table XXIX gives

$$\eta_{xy_{corr}} = \sqrt{\frac{.902^2 - \frac{12-3}{50}}{1 - \frac{12-3}{50}}} = .879$$

and

$$\eta_{yx_{corr}} = \sqrt{\frac{.914^2 - \frac{12-3}{50}}{1 - \frac{12-3}{50}}} = .894.$$

Similarly, for the data in Table XXX the corrected value is .364.

TABLE XXX

COMPUTATION OF THE RATIO OF CORRELATION OF A VARIABLE ON AN ATTRIBUTE

| I. Q. | Type of Book: Adv. & Travel | Biog. & History | Mystery | Romance | Science | Sports | $f_y$ | $y$ | $fy$ | $fy^2$ |
|---|---|---|---|---|---|---|---|---|---|---|
| 140– | 1 | | | | 1 | | 2 | +4 | + 8 | 32 |
| 130– | 2 | 1 | 1 | | | 1 | 5 | +3 | +15 | 45 |
| 120– | 1 | 2 | 2 | 1 | 3 | 2 | 11 | +2 | +22 | 44 |
| 110– | 4 | 3 | 3 | | 2 | 4 | 16 | +1 | +16 | 16 |
| 100– | 8 | 2 | 7 | 2 | 1 | 3 | 23 | 0 | +61 | 0 |
| 90– | 3 | 1 | 7 | 4 | 1 | 5 | 21 | −1 | −21 | 21 |
| 80– | 4 | | 2 | 4 | | 3 | 13 | −2 | −26 | 52 |
| 70– | 1 | | | 1 | | 2 | 4 | −3 | −12 | 36 |
| $f_x$ | 24 | 9 | 22 | 12 | 8 | 20 | 95 | | −59 / +2 | 246 |
| $\Sigma y$ | +2 | +9 | −1 | −13 | +11 | −6 | +22 − 20 \| + 2 | | | |
| $(\Sigma y)^2$ | 4 | 81 | 1 | 169 | 121 | 36 | | | | |
| $\frac{(\Sigma y)^2}{f_x}$ | .17 | 9.00 | .05 | 14.08 | 15.13 | 1.80 | 40.23 | | | |

$$c_y = \frac{2}{95} = .02$$

$$S_y^2 = \frac{246}{95} = 2.5895$$

$$c_y^2 = .0004$$

$$\sigma_y^2 = 2.5891$$

$$\sigma_y = 1.61$$

$$\eta_{yx} = \frac{\sqrt{\frac{40.23}{95} - .0004}}{1.61} = .40$$

Although this is the most commonly used correction formula for this purpose, another is probably preferable. It is

$$\eta_{corr} = \sqrt{\frac{(N-1)\eta^2 - n + 1}{N - n}}.$$

This gives, for Table XXIX,

$$\eta_{xy_{corr}} = \sqrt{\frac{(50-1).902^2 - 12 + 1}{50 - 12}} = .872$$

and

$$\eta_{yx_{corr}} = \sqrt{\frac{(50-1).914^2 - 12 + 1}{50 - 12}} = .888,$$

and, for Table XXX, .336—results not very different from those secured by the other formula.

## The bi-serial coefficient of correlation

Although the ratio of correlation may be found when the attribute concerned is grouped dichotomously, that is, into only two classes, the common measure of relationship between such an attribute and a variable is the *bi-serial coefficient of correlation*. This is a rectilinear measure, ranging from $-1.00$ to $+1.00$, roughly comparable with the ordinary coefficient of correlation between two variables. It is based on the assumption that the attribute involved is in reality a continuous and normally distributed variable which, however, is grouped dichotomously and that its two classes are not extremely unequal in their frequencies, not more so than in the ratio of about 9 to 1. If the distribution of the variable is platykurtic, a result exceeding $\pm 1.00$ may be obtained.

There are several variations of the formula. Probably the easiest to use is

$$\frac{(M_p - M_N)p}{.3989 h_N \sigma_N}.$$

In this $p$ is the fraction of the cases in the larger of the two classes of the attribute, $N$ has its regular meaning as the total number of cases, and $h$ is the height of the normal curve at the distance from the mean such that $p - .5$ of the area is included

between an ordinate erected thereat and the mean. This height is found from a table of the normal curve.

The computation of a bi-serial coefficient is shown in Table XXXI, which contains data comparing passes and failures on

TABLE XXXI

COMPUTATION OF THE BI-SERIAL COEFFICIENT OF CORRELATION

| Test Score | Single Element | | | $d$ | $fd_p$ | $fd_N$ | $fd_N^2$ |
|---|---|---|---|---|---|---|---|
| | Passed | Failed | Total | | | | |
| 110– | 3 | | 3 | +5 | +15 | +15 | 75 |
| 100– | 7 | 2 | 9 | +4 | +28 | +36 | 144 |
| 90– | 12 | 1 | 13 | +3 | +36 | +39 | 117 |
| 80– | 14 | 4 | 18 | +2 | +28 | +36 | 72 |
| 70– | 18 | 6 | 24 | +1 | +18 | +24 | 24 |
| 60– | 22 | 10 | 32 | 0 | +125 | +150 | 0 |
| 50– | 15 | 12 | 27 | −1 | −15 | −27 | 27 |
| 40– | 9 | 15 | 24 | −2 | −18 | −48 | 96 |
| 30– | 6 | 13 | 19 | −3 | −18 | −57 | 171 |
| 20– | 2 | 7 | 9 | −4 | −8 | −36 | 144 |
| Total | 108 | 70 | 178 | | −59 | −168 | 870 |
| | | | | | +66 | −18 | |

$$p = \frac{108}{178} = .607 \qquad h = \text{height including } .607 - .5 \text{ of area} = .964$$

$$c_p = \frac{+66}{108} = +.611 \quad M_p = 65. + .611 \times 10. = 71.11$$

$$c_N = \frac{-18}{178} = -.101 \quad M_N = 65. - .101 \times 10. = 63.99$$

$$S_N^2 = \frac{870}{178} = 4.8876$$

$$c_N^2 = .0102$$

$$\sigma_N^2 = 4.8774$$

$$\sigma_N = 2.208$$

$$i = 10.$$

$$\sigma_N = 22.08$$

$$r_{bis} = \frac{(71.11 - 63.99).607}{.3989 \times .964 \times 22.08} = \frac{4.3218}{8.4906} = +.51$$

one element of a test with total test scores. Since there are more passes than failures, $p$ is the fraction of passes, $\frac{108}{178}$, which equals .607. This minus .5 gives .107, the area between the mean and an ordinate at the point for which $h$ is found. By finding .107 in the third column of either table in the Appendix and interpolating to secure the corresponding entry in the second column, $h$ is determined to be about .964. By the usual method $M_p$ is found to be 71.11, $M_N$ is 63.99, and $\sigma_N$ is 22.08. Substituting these in the formula, we have

$$r_{bis} = \frac{(71.11 - 63.99).607}{.3989 \times .964 \times .22.08} = +.51,$$

the correlation between scores on the single element and those on the whole test.

The bi-serial coefficient is roughly comparable with the product-moment coefficient, but closer comparability may be secured by multiplying the former by $\frac{.3989h}{\sqrt{p(1. - p)}}$. For the data in Table XXXI this multiplier becomes

$$\frac{.3989 \times .964}{\sqrt{.607(1. - .607)}},$$

which equals .787. This times +.51 is +.40, the improved estimate of what $r$ would be for the same data. The same result may be obtained without actually calculating the bi-serial coefficient by the use of the formula

$$r = \frac{(M_p - M_N)\sqrt{p}}{\sigma_N\sqrt{1. - p}}.$$

For these data it gives

$$r = \frac{(71.11 - 63.99)\sqrt{.607}}{22.08\sqrt{1. - .607}} = +.40.$$

The bi-serial coefficient has been used frequently for the purpose of measuring the validity of single test elements. If, as in the example above, score on an element is correlated with that on the whole test, the result is a measure of internal validity; if it is correlated with an outside criterion, the result

is a measure of external validity. There are numerous other methods of measuring element validity, but this is generally considered to be one of the best.

If, as is usual when element validity is being investigated, all that is desired is a comparable index of the validity of elements on the same test, a simplified form of the formula may be employed. Since .3989 and $\sigma_N$ are common to the bi-serial coefficients of all the elements, they may be omitted, thus leaving

$$\frac{(M_p - M_N)p}{h}.$$

Furthermore, if the same assumed means are used, for $p$ and $N$, respectively, throughout, the corrections may be substituted for the means, giving

$$\frac{(c_p - c_N)p}{h}.$$

Results from either of these abbreviated formulas are perfectly correlated with bi-serial coefficients. For the data in Table XXXI, the second gives

$$\frac{[.611 - (-.101)].607}{.964} = .45.$$

This is .88 × +.51, the bi-serial coefficient, and all values of

$$\frac{(c_p - c_N)p}{h}$$

for elements on the same test will be

$$.88r_{bis}$$

for the same elements.

It has been suggested that simplification may go a step farther and $h$ also be dropped. The results tend to arrange elements in the same order of validity, but to indicate differences in their degrees of validity less accurately as differences in $p$ become greater.

## The coefficient of contingency

The best measure of the correlation of attributes, or of an attribute and a variable, when the classification is more than two-fold, is the *coefficient of contingency*, abbreviated $C$ or $CC$.

The formula for this coefficient is given in several forms, some rather complex, but may be stated quite simply as

$$C = \sqrt{\frac{T - 1}{T}}.$$

In this $T$ represents $\Sigma\left(\frac{f_{rc}^2}{f_r f_c}\right)$, in which $f_{rc}$ is the frequency in a given cell of the table, $f_r$ the total frequency in the row in which the cell lies, and $f_c$ that in the column. Thus the effect of the formula is to compare the number of cases in each cell with the number that would fall there by pure chance if each row and column had its stated frequency.

As an example of data for which this method of measuring correlation is suitable, we may employ school marks and major subjects. Table XXXII contains such data for a marking system of five letters and six major fields The formula given above, in which

$$T = \Sigma\left(\frac{f_{rc}^2}{f_r f_c}\right),$$

calls for dividing the square of each entry in Part A, which contains the tabulated data, by the product of the total frequency of its row by that of its column. Thus, for the first cell at the upper left, $4^2$ should be divided by $21 \times 25$; for the second cell of the top row, $3^2$ by $21 \times 12$; and so on. This can be done for each cell and results totalled, giving $T$, but an easier method is shown in Part B. In it each squared entry is divided by the frequency of its column, these quotients summed for each row, and each total divided by the frequency of that row. Thus, for the first row, $4^2$ divided by 25, $3^2$ divided by 12, and so on, are summed and their total divided by 21, which gives .1673. The same procedure is applied to each row, and the results added, their sum being 1.0430. This is $T$; hence

$$C = \sqrt{\frac{1.0430 - 1}{1.0430}} = .20.$$

Instead of summing by rows, we may do the same by columns. For this the square of each entry is divided by the frequency of its row, these quotients summed by columns, and

TABLE XXXII

COMPUTATION OF THE COEFFICIENT OF CONTINGENCY

PART A

| Mark | Major Subject | | | | | | |
|---|---|---|---|---|---|---|---|
| | English | For. Lang. | H. E. & I. A. | Math. | Science | Soc. Stud. | Total |
| *A* | 4 | 3 | 6 | 2 | 3 | 3 | 21 |
| *B* | 8 | 4 | 13 | 6 | 8 | 6 | 45 |
| *C* | 7 | 4 | 10 | 7 | 8 | 5 | 41 |
| *D* | 5 | 1 | 4 | 5 | 3 | 3 | 21 |
| *E* | 1 | 0 | 1 | 1 | 2 | 1 | 6 |
| Total | 25 | 12 | 34 | 21 | 24 | 18 | 134 |

PART B

$$\frac{1}{21}\left(\frac{16}{25}+\frac{9}{12}+\frac{36}{34}+\frac{4}{21}+\frac{9}{24}+\frac{9}{18}\right) = .1673$$
$$\frac{1}{45}\left(\frac{64}{25}+\frac{16}{12}+\frac{169}{34}+\frac{36}{21}+\frac{64}{24}+\frac{36}{18}\right) = .3388$$
$$\frac{1}{41}\left(\frac{49}{25}+\frac{16}{12}+\frac{100}{34}+\frac{49}{21}+\frac{64}{24}+\frac{25}{18}\right) = .3079$$
$$\frac{1}{21}\left(\frac{25}{25}+\frac{1}{12}+\frac{16}{34}+\frac{25}{21}+\frac{9}{24}+\frac{9}{18}\right) = .1724$$
$$\frac{1}{6}\left(\frac{1}{25}+\frac{0}{12}+\frac{1}{34}+\frac{1}{21}+\frac{4}{24}+\frac{1}{18}\right) = .0566$$
$$\overline{1.0430}$$

$$C = \sqrt{\frac{1.0430-1}{1.0430}} = .20 \qquad C_{corr} = \frac{.20}{\sqrt[4]{\dfrac{(5-1)(6-1)}{5\times 6}}} = .22$$

each total divided by the column frequency. The first column, for example, gives

$$\frac{1}{25}\left(\frac{16}{21}+\frac{64}{45}+\frac{49}{41}+\frac{25}{21}+\frac{1}{6}\right) = .1895.$$

The sum of these results is $T$, as before. In this case it is 1.0429, the slight difference from 1.0430 being due to dropping

decimals. It is advisable to sum in both directions, for checking.

The coefficient of contingency may have values from .00 up to

$$\sqrt[4]{\frac{(n_1 - 1)(n_2 - 1)}{n_1 n_2}},$$

in which $n_1$ is the number of classes in one direction and $n_2$ that in the other. In this example they are 5 and 6; hence $C$ could not exceed

$$\sqrt[4]{\frac{(5 - 1)(6 - 1)}{5 \times 6}} = .90.$$

If the obtained value is divided by the upper limit, the result—which may be symbolized by $C_{corr}$—is roughly equivalent to $r$. Here

$$C_{corr} = .20 \div .90 = .22.$$

In order that the coefficient of contingency and the ratio of correlation may be compared directly, the former has been computed for the data in Table XXX, for which $\eta_{yx} = .40$. For them, $C = .52$. Conversely, for those in Table XXXII, $\eta_{yx} = .17$.

A significant value of $C$ merely indicates that there is some relationship between the two series of data, but gives no information as to its direction. Rows or columns of a contingency table may be interchanged among themselves without affecting the value of the obtained coefficient.

## Product-moment correlation of attributes

If, in cases involving correlation between an attribute and a variable or two attributes, the attribute or attributes concerned are of such a nature that they may legitimately be considered to form a normal distribution, their product-moment coefficient of correlation may be found. All that is necessary to make this possible is to assign numerical values to the classes according to the method explained later, in Chapter XIII. After this has been done, the procedure for comput-

ing $r$ is as usual except that care must be exercised to employ proper $d$ values.

## Tetrachoric correlation

The best type of correlation to employ when both variables or attributes are dichotomous—that is, divided into only two classes—is *tetrachoric correlation.* It is valid only to the degree that the variables or attributes would be distributed normally if divided into more classes. The number of cases should not be quite small, nor should either of the two classes of either variable or attribute contain less than .05 of its total frequency. The computation of tetrachoric coefficients by means of a set of diagrams prepared by Thurstone and others[2] is relatively easy.

Although the computation of tetrachoric $r$ without the use of the diagrams is more difficult, it seems well to illustrate it. Part A of Table XXXIII presents data as to the numbers of

TABLE XXXIII

COMPUTATION OF THE TETRACHORIC COEFFICIENT OF CORRELATION

PART A

| Element No. 2 | Element No. 1 | | |
|---|---|---|---|
| | $R$ | $W$ | $T$ |
| $W$ | 67 | 63 | 130 |
| $R$ | 82 | 38 | 120 |
| $T$ | 149 | 101 | 250 |

PART B

$$p_1 = \frac{149}{250} = .596 \qquad h_1 = .971 \qquad x_1 = .243$$

$$p_2 = \frac{130}{250} = .520 \qquad h_2 = .998 \qquad x_2 = .050$$

$$\frac{.243 \times .050}{2} r^2_{tet} + r_{tet} + \frac{63 \times 82 - 67 \times 38}{.1591 \times 250.^2 \times .971 \times .998} = 0$$

$$.006075 r^2_{tet} + r_{tet} + .2719 = 0$$

$$r_{tet} = \frac{-1 \pm \sqrt{1^2 - 4 \times .006075 \times .2719}}{2 \times .006075} = -.27$$

right and wrong responses on two test elements. It shows that, of the 130 individuals who responded incorrectly to element 2, 67 answered element 1 correctly and 63 incorrectly, and, of the 120 who had element 2 correct, 82 had correct answers to

[2] Chesire, Leona, Saffir, Milton, and Thurstone, L. L., *Computing Diagrams for the Tetrachoric Correlation Coefficient.* Chicago: University of Chicago Bookstore, 1933.

element 1 and 38 incorrect ones. As here, the upper row and the column at the left should represent the larger of each pair of classes. Part B shows the computation. The four cells in the table in Part A are thought of as lettered $a$, $b$, $c$, and $d$, as shown at the right. Thus here, $a = 67$, $b = 63$, $c = 82$, and $d = 38$. The first step is to find $p$—the fraction of the group in the larger class—for each of the two things correlated. Here $p_1 = \frac{149}{250} = .596$ and $p_2 = \frac{130}{250} = .520$. Next, the height of the normal curve at the point that includes $p - .5$ of the area between it and the mean is found for each. For $p_1$ the included area is $.596 - .5 = .096$ and the corresponding height is .971, and for $p_2$ the area is $.520 - .5 = .020$ and the height .998. Finally the standard deviation distance from the mean to the same point—the one including the given area—is found for each. It is usually symbolized by $x$. In this case $x_1 = .243$ and $x_2 = .050$.

| | |
|---|---|
| $a$ | $b$ |
| $c$ | $d$ |

When these quantities have been found, they are inserted in the equation

$$\frac{x_1x_2}{2}r_{tet}^2 + r_{tet} + \frac{bc - ad}{.1591N^2h_1h_2} = 0.$$

In the example, this gives

$$\frac{.243 \times .050}{2}r_{tet}^2 + r_{tet} + \frac{63 \times 82 - 67 \times 38}{.1591 \times 250.^2 \times .971 \times .998} = 0,$$

which reduces to

$$.006075r_{tet}^2 + r_{tet} + .2719 = 0.$$

This may be solved for $r_{tet}$ by any of the standard methods for quadratic equations. Perhaps the most convenient in this case is that based upon the fact that for the standard form of quadratic equation, $ax^2 + bx + c = 0$, the roots may be obtained from

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.$$

Using $r_{tet}$ as $x$, $a$ as .006075, $b$ as 1., and $c$ as .2719, and substituting gives

$$r_{tet} = \frac{-1 \pm \sqrt{1^2 - 4 \times .006075 \times .2719}}{2 \times .006075} = -.27 \text{ or } -164.$$

Since the latter is greater than 1.00, it is impossible; hence $-.27$ is the desired value. It is the coefficient between $p_1$ and $p_2$—that is, between rights on element 1 and wrongs on element 2. Therefore that between rights on the two elements, or wrongs on both, is $+.27$.

Tetrachoric $r$ has the same range of possible value, from $-1.00$ to $+1.00$, as product-moment $r$, and in general may be considered as comparable with it and interpreted in the same way.

## The phi coefficient

For a four-fold table containing data of such nature that each series can be divided into only two classes and cannot be thought of as forming a normal distribution if more finely divided, the $\phi$ (*phi*) *coefficient* is available. To compute it one should arrange the data as for tetrachoric correlation and employ the same four letters, $a$, $b$, $c$, and $d$, to represent the frequencies in the four cells. Also, as elsewhere, $p_1$ is used for the larger of the two fractions into which one series of data is divided and $p_2$ for that of the other. The formula is

$$\phi = \frac{ad - bc}{N^2\sqrt{p_1(1 - p_1)p_2(1 - p_2)}}.$$

If, instead of numbers of cases in the four cells of a table, fractions of the total number of cases are used, with $\alpha$ (alpha) $= \frac{a}{N}$, $\beta$ (beta) $= \frac{b}{N}$, $\gamma$ (gamma) $= \frac{c}{N}$, and $\delta$ (delta) $= \frac{d}{N}$, its numerator becomes $\alpha\delta - \beta\gamma$ and its denominator loses the term $N^2$.

To illustrate the computation of $\phi$, we may assume that the data in Table XXXIII, employed for tetrachoric $r$, are such that $\phi$ rather than $r_{te}$ would be appropriate. We had $p_1 = .596$ and $p_2 = .520$; whence $1 - p_1 = .404$ and $1 - p_2 = .480$. Employing these values, the first formula gives

$$\frac{67 \times 38 - 63 \times 82}{250^2\sqrt{.596 \times .404 \times .520 \times .480}}$$

and the second,

$$\frac{.268 \times .152 - .252 \times .328}{\sqrt{.596 \times .404 \times .520 \times .480}},$$

both of which equal $-.17$.

The $\phi$ coefficient may also be found for tables such as those for which a coefficient of contingency is more commonly determined, but since the latter is a better measure this is rarely done. The two are closely connected, since

$$\phi = \frac{C}{\sqrt{1 - C^2}} \text{ and } C = \sqrt{\frac{\phi^2}{1 - \phi^2}}.$$

Also, $\phi = \sqrt{T - 1}$. For the data in Table XXXII, $\phi = \sqrt{1.0430 - 1.} = .21$, very close to the value already obtained for $C$, .20. Unless both are .00, $\phi$ exceeds $C$ for the same data, the difference becoming larger, the higher their values.

When secured for data for which it is appropriate, $\phi$ may be considered as equivalent to product-moment $r$. If it has been found for continuous variables grouped in two classes each—a practice not recommended—dividing it by .637 renders it more nearly equivalent to $r$ than it would be otherwise.

## EXERCISES AND PROBLEMS

**1.** Compute the ratio of correlation for each part of Exercises 2 and 3 on pages 134 and 135.

**2.** Compute the ratio of correlation for each of the following sets of data:

(*a*)

| Mark | Boys | Girls |
|---|---|---|
| A | 4 | 7 |
| B | 14 | 15 |
| C | 22 | 28 |
| D | 15 | 14 |
| E | 3 | 6 |
| F | 2 | 1 |

(*b*)

| Salary | H. S. Grad. | Non-H. S. Grad. |
|---|---|---|
| $2400– | 2 | 0 |
| 2200– | 1 | 0 |
| 2000– | 4 | 1 |
| 1800– | 5 | 0 |
| 1600– | 11 | 6 |
| 1400– | 19 | 17 |
| 1200– | 14 | 23 |
| 1000– | 8 | 18 |
| 800– | 3 | 7 |
| 600– | 0 | 2 |

(*c*)

| Age | Passed | Failed |
|---|---|---|
| 9–6– | 1 | 1 |
| 9–0– | 0 | 1 |
| 8–6– | 2 | 3 |
| 8–0– | 1 | 2 |
| 7–6– | 3 | 0 |
| 7–0– | 7 | 3 |
| 6–6– | 23 | 4 |
| 6–0– | 29 | 4 |
| 5–6– | 4 | 0 |
| 5–0– | 2 | 0 |

**3.** Compute the bi-serial coefficient of correlation for Parts (*a*) and (*b*) in Exercise 2.

**4.** Compute the coefficient of contingency for each of the following sets of data:

(*a*)

| Intel. Group | Place of Birth | | | |
|---|---|---|---|---|
| | Farm | Small Town | Small City | Large City |
| Sup. | 2 | 1 | 3 | 7 |
| Good | 4 | 5 | 6 | 13 |
| Aver. | 12 | 14 | 16 | 23 |
| Poor | 10 | 16 | 14 | 11 |
| Inf. | 3 | 6 | 5 | 8 |

(*b*)

| Type of School | Years of Training | | | | |
|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 |
| Rural | 24 | 62 | 28 | 16 | 3 |
| Town Elem. | 12 | 85 | 31 | 22 | 10 |
| City Elem. | 8 | 33 | 32 | 54 | 13 |
| Town High | 0 | 0 | 6 | 38 | 16 |
| City High | 0 | 0 | 1 | 30 | 49 |

(*c*)

| Best Subject | Intended Vocation | | | | |
|---|---|---|---|---|---|
| | Prof. | Semi-Prof. | Sk. Tr. | Unsk. Tr. | Uncer. |
| Lang. | 4 | 3 | 1 | 2 | 6 |
| Math. | 11 | 14 | 5 | 3 | 5 |
| Sci. | 9 | 13 | 10 | 4 | 3 |
| Hist. | 6 | 4 | 2 | 0 | 7 |
| Eng. | 5 | 4 | 2 | 2 | 2 |
| I. A. | 4 | 6 | 12 | 3 | 3 |
| H. E. | 1 | 3 | 7 | 2 | 6 |

**5.** Correct each coefficient of contingency found in Exercise 4 by the formula given in the text.

**6.** Compute the tetrachoric coefficient of correlation for each of the sets of data given below.

(*a*)

| French | English | |
|---|---|---|
| | Passed | Failed |
| Passed | 92 | 6 |
| Failed | 16 | 12 |

(*b*)

| | Successful | Failing |
|---|---|---|
| Grad. work | 25 | 4 |
| No grad. work | 45 | 10 |

(*c*)

| | High | Low |
|---|---|---|
| Nonparticipating | 71 | 35 |
| Participating | 22 | 32 |

**7.** Compute the phi coefficient for each set of data below:

(*a*)

| | Elementary | High |
|---|---|---|
| Men | 16 | 14 |
| Women | 95 | 31 |

(*b*)

| | Non-College | College |
|---|---|---|
| Public | 1436 | 387 |
| Private | 241 | 564 |

(*c*)

| French | Latin | |
|---|---|---|
| | Not Taken | Taken |
| Not taken | 147 | 73 |
| Taken | 64 | 35 |

## REFERENCES

Ezekiel, Mordecai, *Methods of Correlation Analysis.* Second Edition. Chap. XVII; pages 152–158. New York: John Wiley & Sons, Inc., 1941.

Garrett, Henry E., *Statistics in Psychology and Education.* Second Edition. Pages 366–377, 387–402. New York: Longmans, Green & Company, Inc., 1937.

Guilford, J. P., *Fundamental Statistics in Psychology and Education.* Pages 231–254. New York: McGraw-Hill Book Company, Inc., 1942.

Kenney, John F., *Mathematics of Statistics.* Part I, pages 198–207. New York: D. Van Nostrand Company, Inc., 1939.

Lindquist, E. F., *Statistical Analysis in Educational Research.* Pages 235–246. Boston: Houghton Mifflin Company, 1940.

Mode, Elmer B., *The Elements of Statistics.* Pages 303–307. New York: Prentice-Hall, Inc., 1941.

Monroe, Walter S., Editor, *Encyclopedia of Educational Research.* Pages 370–371. New York: The Macmillan Company, 1941.

Peters, Charles C., and Van Voorhis, Walter R., *Statistical Procedures and Their Mathematical Bases.* Chap. XIII; pages 312–324. New York: McGraw-Hill Book Company, Inc., 1940.

# CHAPTER XIII

# Uses of the Normal Probability Curve

## Determining fraction of cases between two points

It is frequently useful to be able to determine the fraction of cases falling between two points of a normal distribution or in each of a number of divisions into which such a distribution may be divided. To do the former, all that is necessary is to find the standard or median deviation distances of the two points from the mean, look up in the Appendix the area included between each such distance and the mean, and take the difference of the areas. For an example, we may refer back to the data in Table XI on page 95. Let us assume that we wish to know what fraction, or how many, of the cases would fall between 82 and 90 if the distribution were normal. The mean is 70.25 and the standard deviation 11.47. Therefore 82 is $\frac{82. - 70.25}{11.47} = 1.02\sigma$ above the mean; 90 is $\frac{90. - 70.25}{11.47} =$ $1.72\sigma$ above it. From the first table in the Appendix we find that approximately .346 of the area lies between the mean and a point $1.02\sigma$ from it, and .457 of it between the mean and a point $1.72\sigma$ from it. Therefore the fraction of cases between 82 and 90, if normality is assumed, is $.457 - .346 = .111$. The number of cases so located is $.111 \times 110$, the total number in Table XI, or approximately 12.

Since a very common situation in which the areas of parts of the normal surface are wanted is the assignment of marks, the procedure therefor will be illustrated. There has been much argument as to the application of the normal, or any other, curve to marking practice and the writer does not wish to be understood as recommending rigid adherence to any pre-

determined distribution, particularly in small or selected groups. He does believe, however, that such distributions have a place as at least general guides and that for typical unselected groups of pupils the normal is the best one to follow.

Several decisions must be made before the actual determination of the fraction of pupils to receive each mark begins. Probably the first is the question of how many marks to employ. Any number may be chosen, but the writer recommends from five to seven. Five—*A*, *B*, *C*, *D*, and *E*—will be used here. The next decision is as to whether the normal surface is to be divided into portions with equal base lines or not. Ordinarily it is assumed that it is so divided; hence that assumption will be made in this instance. Finally, since the normal surface extends to infinity, a decision must be made as to where to cut it off so as to permit equal division of the base line. For the purpose of assigning marks, it is usually cut off at $\pm 2.5\sigma$ or $\pm 3.0\sigma$ from the mean. In this case the former, $\pm 2.5\sigma$, will be used.

Figure 21 illustrates a normal curve and its underlying sur-

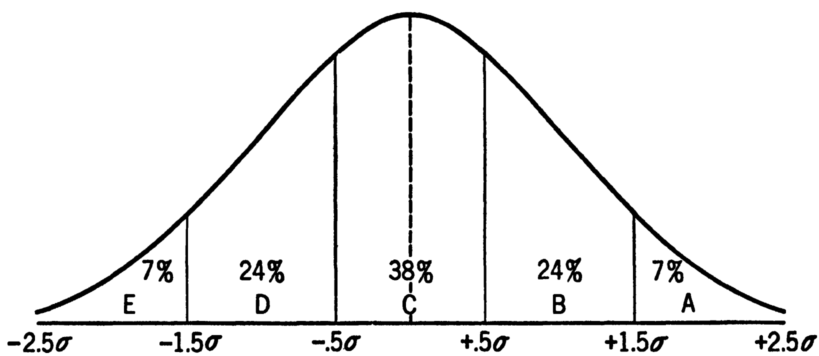



**Fig. 21. Division of Normal Surface Into Five Parts with Equal Base Lines**

face divided as has just been described. Since it extends from $-2.5\sigma$ to $+2.5\sigma$, its total width is $5.0\sigma$; hence each of the five divisions has a base line $1.0\sigma$ in length. Therefore the points separating the divisions are, from left to right, $-1.5\sigma$, $-.5\sigma$, $+.5\sigma$, and $+1.5\sigma$. In order to include the whole area,

the small portions thereof beyond the extreme points—here $-2.5\sigma$ and $+2.5\sigma$—are included in the extreme divisions. By means of the standard deviation table in the Appendix we find that .0668 of the area under the curve is beyond $1.5\sigma$ from the mean; hence that fraction—or about 7 per cent—of marks should be *E*'s and as many *A*'s. Since .3085 of the area is beyond $.5\sigma$ and this minus .0668 is .2417, that fraction—or about 24 per cent—of marks should be *D*'s and as many *B*'s. Finally, the area from $-.5\sigma$ to $+.5\sigma$ is twice .1915—or .3830—of the whole; so that about 38 per cent of the marks should be *C*'s. The 7,24,38,24,7 distribution thus derived is the most commonly recommended and used normal distribution of marks.

If the points at which the normal surface is cut off before dividing are closer to the mean, the fractions in the central divisions are smaller and those in the extreme ones greater; if the points are farther from the mean, the fractions in the central divisions are greater and those in the extreme ones smaller. For example, if $\pm 3.0\sigma$ instead of $\pm 2.5\sigma$ is taken as the limit, the per cents of area are approximately $3\frac{1}{2}$, 24, 45, 24, and $3\frac{1}{2}$.

It is not necessary that the base-line distances of the various portions of area be equal. The area under the curve, and correspondingly the per cents of marks, may follow any arbitrary pattern of division desired. If a group is of above-average ability and is doing work of corresponding quality, a distribution with larger per cents of high marks and smaller ones of low marks may be adopted, and vice versa. For example, in such a case of a superior group the per cents of marks considered ideal may be set as follows: $E - 5$, $D - 15$, $C - 40$, $B - 30$, $A - 10$.

The *standard, sigma,* or *z-score* method of assigning marks is based upon an ideal distribution of marks, which may be normal or any other type desired. The limits, in terms of *standard, $\sigma$,* or *z-scores,* of the divisions of the ideal distribution are found and then marks are assigned according to the limits between which scores fall, regardless of what fraction of marks actually falls within any division. Thus this method assumes an ideal distribution and assumes that the mean of the actual distribu-

tion is the same as the mean of the ideal one, but assigns marks according to the actual distribution of scores.

To illustrate this procedure, the data in Table IX on page 89 may be used. Let us assume that we wish to change the 15 scores listed there into marks and that the 7,24,38,24,7 distribution is to be considered the ideal one. First we need to find the four points, in terms of $\sigma$ distances from the mean, which divide the normal surface into such parts. These, as already given earlier in this section, are $-1.5\sigma$, $-.5\sigma$, $+.5\sigma$, and $+1.5\sigma$. Using the mean, 41.8, and the standard deviation, 31.55, found in Table IX, we compute the four division points as follows:

$$M - 1.5\sigma = 41.8 - 1.5 \times 31.55 = -5.5$$
$$M - .5\sigma = 41.8 - .5 \times 31.55 = 26.0$$
$$M + .5\sigma = 41.8 + .5 \times 31.55 = 57.6$$
$$M + 1.5\sigma = 41.8 + 1.5 \times 31.55 = 89.1.$$

Therefore marks should be assigned as follows:

*E* below −5.5
*D* from −5.5 to 26.0
*C* from 26.0 to 57.6
*B* from 57.6 to 89.1
*A* above 89.1.

According to this, one score, 98, becomes *A*; four, 81,81,72, and 85, become *B*'s; two, 52 and 37, *C*'s; eight, 18,13,15,15,23,12,14, and 11, *D*'s; none is an *E*.

If, instead of the 7,24,38,24,7 distribution, another—such as the 5,15,40,30,10 one, for example—is preferred, the only difference is that the division points therefor are employed. From the Appendix they are found to be in order, $-1.65\sigma$, $-.84\sigma$, $+.25\sigma$, and $+1.28\sigma$. In terms of scores these are:

$$M - 1.65\sigma = 41.8 - 1.65 \times 31.55 = -10.3$$
$$M - .84\sigma = 41.8 - .84 \times 31.55 = 15.3$$
$$M + .25\sigma = 41.8 + .25 \times 31.55 = 49.7$$
$$M + 1.28\sigma = 41.8 + 1.28 \times 31.55 = 82.2.$$

Accordingly, marks should be given thus:

*E* below −10.3
*D* from −10 3 to 15.3
*C* from 15.3 to 49.7
*B* from 49.7 to 82.2
*A* above 82.2.

This gives two *A*'s, 98 and 85; four *B*'s, 81,52,81, and 72; three *C*'s, 18,23, and 37; six *D*'s, 13,15,15,12,14, and 11; and again no *E*'s.

## Changing qualitative data into a normal distribution

It is frequently convenient to change nonquantitative data into numerical form. If they are qualitative—that is, such that they can be arranged in order of degrees of merit or excellence—they can be transmuted into a quantitative or numerical distribution. Naturally the form of this distribution should be such as fits the characteristic or fact involved. Since in many cases the appropriate form is the normal distribution, the procedure for transmuting qualitative into quantitative data according to that form will be explained.

As data for the purpose just stated, we may take ratings of teachers, expressed verbally. Let us suppose, therefore, that 80 teachers have been rated as follows: *Excellent*—10; *superior*—25; *average*—27; *inferior*—14; *very poor*—4; and that it is assumed they should be normally distributed. The general procedure is to express each of the five ratings in terms of the average distance of the portion of area under a normal curve that represents it from the mean or center of the curve. This distance may be in terms of any measure of variability, but the standard deviation is most commonly employed; hence, the distance used is a standard or $z$-score.

Figure 22 illustrates the division of the normal curve into

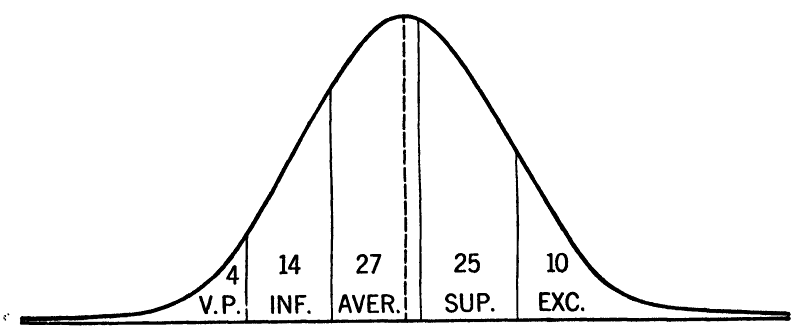



**Fig. 22. Graphical Representation of Data in Table XXXIV**

portions corresponding to the five ratings as given in the last paragraph, and Table XXXIV contains the computation necessary to determine their standard deviation distance from the

TABLE XXXIV

TRANSMUTATION OF QUALITATIVE DATA INTO STANDARD DEVIATION DISTANCES FROM THE MEAN OF A NORMAL DISTRIBUTION

| Rating | Number of Cases | Fraction of Area | Fraction of Area Between Outer Ordinate and Maximum Ordinate | Height of Ordinate | Difference in Heights of Ordinates | Mean Standard Deviation Distance from Mean |
|---|---|---|---|---|---|---|
| | | | .5000 | .0000 | | |
| *Excellent* | 10 | .1250 | | | +.5160 | +1.65 |
| | | | .3750 | .5160 | | |
| *Superior* | 25 | .3125 | | | +.4717 | +.60 |
| | | | .0625 | .9877 | | |
| *Average* | 27 | .3375 | | | −.2359 | −.28 |
| | | | .2750 | .7518 | | |
| *Inferior* | 14 | .1750 | | | −.4933 | −1.12 |
| | | | .4500 | .2585 | | |
| *Very Poor* | 4 | .0500 | | | −.2585 | −2.06 |
| | | | .5000 | .0000 | | |

mean. Thus, in the figure, the portion of the area at the right is $\frac{10}{80}$ or .125 of all; the next is $\frac{25}{80}$ or .3125 of all; and so on.

In the table the first two columns contain the ratings and the number of each, and the third, the latter divided by the total number, 80. The fourth-column entries are the fractions of the whole area under the curve between the ordinate at the outer limit of each of the five portions and the mean or maximum ordinate. Since the outer ordinate of each extreme portion is at infinity from the mean ordinate, half or .5000 of the area is between it and the latter; since .1250 of the area is in the *excellent* division, .5000 − .1250 = .3750 of the area is between the outer ordinate of the next or *superior* division and the mean ordinate; and so on. The heights of the ordinates that include the areas just given are to be found in the fifth column. At infinity, corresponding to an area of .5000, the height is .0000; for each of the other fractions of area the height of the outer ordinate which includes it is found from the Appendix. Thus the ordinate that includes an area of .3750

between it and the mean ordinate is .5160 high; that which includes an area of .0625 is .9877 high; and so on. The next column contains the differences in heights of the ordinates just found, each difference having the sign given by subtracting a height from the one just below it. Thus .5160 − .0000 = +.5160; .9877 − .5160 = +.4717; and so on. Finally, the desired standard deviation distance of each group of ratings from the mean is found by dividing each difference by 2.5066 times the corresponding fraction of area. For the *excellent* group, $\frac{+.5160}{2.5066 \times .1250} = +1.65$, which is its average $\sigma$ distance from the mean. Similarly, the average distance of the *superior* group is found to be $+.60\sigma$, that of the *average* group $-.28\sigma$, that of the *inferior* group $-1.12\sigma$, and that of the *very poor* group $-2.06\sigma$. These may then be used where numerical rather than verbal ratings are desired.

It is possible, by a somewhat easier method, to secure median instead of mean distances, but for most purposes they are less satisfactory. They are regularly somewhat less, positively or negatively, than the mean distances. To secure them, all that is necessary is to find the fraction of area between the ordinate that divides each portion into two equal parts and the mean and find its distance from the mean. Thus, for the *excellent* group, .4375 of the area lies between the dividing point and the mean, and the corresponding $\sigma$ distance is +1.53. For the other ratings the median $\sigma$ distances are, respectively, +.58, −.27, −1.09, and −1.96.

Probably the most frequent occasion for the application of this procedure is in cases wherein ratings or other data from several sources are to be combined or compared. For example, if one supervisor assigns ratings to a group of teachers according to the distribution in Table XXXIV, whereas another, rating the same 80 teachers, gives 5 of them *excellent*, 12 *superior*, 33 *average*, 19 *inferior*, and 11 *very poor*, it is evident that the latter's standards are more severe than those of the former Therefore an *excellent* rating as given by him represents a higher degree of merit than a similar rating by the first rater. Similarly, the significance of the other four depends upon the dis-

tribution of ratings. To combine or compare them, we should express the meaning of those given by each rater in similar fashion.

To illustrate further the application of this procedure, we may suppose that each of the 80 teachers was rated by two supervisors, A and B, who gave the two sets of ratings reported above and also by two others, C and D, whose ratings along with those of A and B are given in Part A of Table XXXV. From the data there it is evident that C gave more extreme ratings, both high and low, and D fewer extreme ones, than A

TABLE XXXV

RATINGS ASSIGNED EIGHTY TEACHERS BY FOUR SUPERVISORS AND STANDARD DEVIATION DISTANCES FROM THE MEAN CORRESPONDING THERETO

PART A. ASSIGNED RATINGS

| Rating | Supervisor | | | |
|---|---|---|---|---|
| | *A* | *B* | *C* | *D* |
| *Excellent* | 10 | 5 | 10 | 4 |
| *Superior* | 25 | 12 | 18 | 14 |
| *Average* | 27 | 33 | 22 | 42 |
| *Inferior* | 14 | 19 | 19 | 17 |
| *Very Poor* | 4 | 11 | 11 | 3 |

PART B. STANDARD DEVIATION DISTANCES

| Rating | Supervisor | | | |
|---|---|---|---|---|
| | *A* | *B* | *C* | *D* |
| *Excellent* | +1.65 | +1.97 | +1.65 | +2.06 |
| *Superior* | +.60 | +1.11 | +.73 | +1.12 |
| *Average* | −.28 | +.22 | −.03 | +.03 |
| *Inferior* | −1.12 | −.67 | −.67 | −1.11 |
| *Very Poor* | −2.06 | −1.60 | −1.60 | −2.18 |

and B. In Part B of Table XXXV the mean $\sigma$ distance from the mean, or $z$-score, corresponding to each rating by each supervisor is given. Those for A are as found in Table XXXIV and the others were computed by the same process. To secure the mean rating given a teacher by the four supervisors, the four $\sigma$ distances corresponding to the four ratings should be averaged. Thus, if Miss Smith was rated *superior*, *average*, *superior*, and *superior* by the four supervisors, in the order used in the tables, her mean rating is +.67, the average of +.60, +.22, +.73, and +1.12. Similarly, if Miss Jones is rated *average*, *inferior*, *inferior*, and *average*, in order, her mean rating is $\frac{-.28 - .67 - .67 + .03}{4} = -.40$.

## Comparing an actual with an ideal distribution

Workers with educational data sometimes wish to compare actual with ideal or hypothetical distributions. Since the most frequently used type of ideal distribution is the normal, the procedure will be illustrated for that type, but it may be applied to others also. Generally the simplest method of making such a comparison is to graph the actual and the theoretical distributions on the same figure and scale. An example of so doing is shown in Figure 23. The solid line therein represents the actual distribution; the broken line, the ideal or normal one. The latter was constructed by placing its center at the mean, where its height is given by the formula $y_0 = \frac{N}{\sigma\sqrt{2\pi}} = \frac{N}{2.5066\sigma}$,[1] and then, by means of standard deviation distances and the first table in the Appendix, locating points through which it was drawn. Here $y_0$, the maximum ordinate, $= \frac{750.}{2.5066 \times 1.864} =$ 160.50. By laying off distances $.25\sigma$ or even $.5\sigma$ apart in each direction from 64.80, the mean, and multiplying 160.50 by the entries in the height column corresponding to the $\sigma$ values employed, the points through which the best-fitting normal

---

[1] If the data are grouped in classes and the figure constructed on that basis, as here, $\sigma$ should be in terms of intervals; if they are not grouped, $\sigma$ should be in terms of score units.

curve should pass were determined. Thus, for example, $.5\sigma = .5 \times 9.32 = 4.66$. This distance below the mean gives $64.80 - 4.66 = 60.14$ and above the mean $64.80 + 4.66 = 69.46$. The height entry corresponding to $.5\sigma$ is .8825; hence the height of the desired normal curve at 60.14 and 69.46 is $.8825 \times 160.50 = 141.64$. Similarly, at $64.80 - 9.32 = 55.48$ and at $64.80 + 9.32 = 74.12$ its height is $.6065 \times 160.50 = 97.34$, and so on.

If the theoretical frequencies given in the sixth column of Table XXXVI have been determined, the best-fitting normal

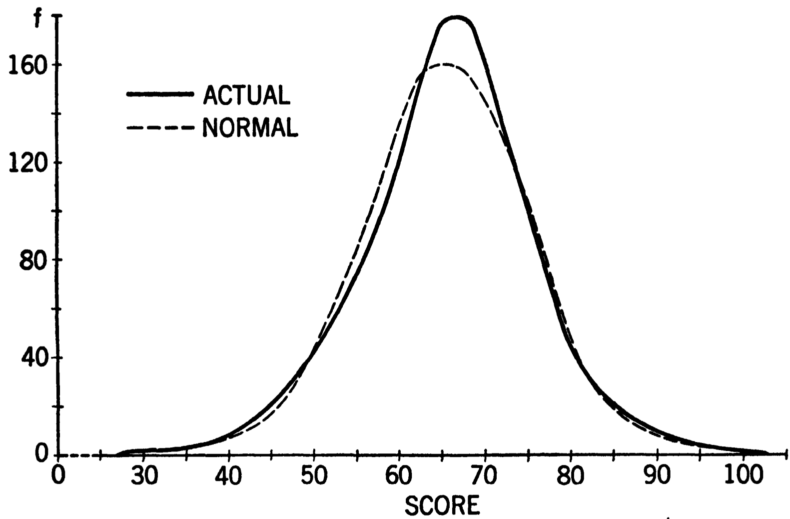



**Fig. 23. Actual and Best-Fitting Normal Curve for Same Data**

curve can be constructed more easily from them than by the method used above. The entries in this column are the heights of the normal curve at the mid-points of the classes; so all that is necessary is to locate points in accordance with that fact and draw a curve through them. The first such point is at a height of .41 above 32.5, the mid-point of the 30-class; the next is 2.21 above 37.5; the next is 9.24 above 42.5; and so on.

Inspection of Figure 23 reveals that no very precise idea of the amount of difference, or of the chances that it is significant, can be obtained therefrom. For such purposes computational rather than graphic methods are needed.

TABLE XXXVI

COMPARISON OF ACTUAL AND NORMAL CURVES

| Lower Limit | Actual Freq. | Mid-Point | Distance from $M$ | $\sigma$ Distance | Theor. Freq. | Difference | $\frac{\text{Diff.}^2}{\text{Theor. } f}$ | $\epsilon_{\text{diff.}}$ | $\frac{\text{Diff.}}{\epsilon_{\text{diff.}}}$ | Chances to 1 |
|---|---|---|---|---|---|---|---|---|---|---|
| 95– | 1 | 97.5 | 32.7 | 3.51 | .34 | | | | | |
| 90– | 3 | 92.5 | 27.7 | 2.97 | 1.96 | 2.48 | .59 | 3.22 | .77 | 3.5 |
| 85– | 9 | 87.5 | 22.7 | 2.44 | 8.22 | | | | | |
| 80– | 22 | 82.5 | 17.7 | 1.90 | 26.40 | 4.40 | .73 | 5.05 | .87 | 4.2 |
| 75– | 54 | 77.5 | 12.7 | 1.36 | 63.72 | 9.72 | 1.48 | 7.66 | 1.27 | 8.8 |
| 70– | 115 | 72.5 | 7.7 | .83 | 113.70 | 1.30 | .01 | 9.82 | .13 | 1.2 |
| 65– | 178 | 67.5 | 2.7 | .29 | 153.82 | 24.18 | 3.80 | 11.06 | 2.19 | 69.4 |
| 60– | 156 | 62.5 | 2.3 | .25 | 155.38 | .62 | .00 | 11.10 | .06 | 1.1 |
| 55– | 105 | 57.5 | 7.3 | .78 | 118.36 | 13.36 | 1.50 | 9.98 | 1.31 | 10.1 |
| 50– | 63 | 52.5 | 12.3 | 1.32 | 67.21 | 4.21 | .26 | 7.83 | .51 | 2.4 |
| 45– | 30 | 47.5 | 17.3 | 1.86 | 28.55 | 1.45 | .07 | 5.24 | .23 | 1.6 |
| 40– | 10 | 42.5 | 22.3 | 2.39 | 9.24 | | | | | |
| 35– | 2 | 37.5 | 27.3 | 2.93 | 2.21 | 2.14 | .39 | 3.42 | .63 | 2.8 |
| 30– | 2 | 32.5 | 32.3 | 3.47 | .41 | | | | | |
| | 750 | | | | 749.52 | | 8.83 | | | |

$N = 750$ $M = 64.80$ $\sigma = 9.32$ $y_0 = 160.50$ $\chi^2 = 8.83$ $P = .46$

There are several computational methods of making comparisons between actual and theoretical curves, not all of which lead to the same conclusion in a given instance. The most commonly used one is that known as the $\chi^2$ (*Chi-square*) *Test for Goodness of Fit.* Its validity depends upon several assumptions, and interpretations based upon it are only approximate. It should be based upon a sample of not less than 500 cases, with not less than 10, or perhaps 5, in any single class. When classes contain fewer than this number, two or more adjacent ones should be combined for the last steps in the computation.

The method just mentioned is illustrated in Table XXXVI. The first two columns thereof contain the actual distribution; the third, the class mid-points; and the fourth, the distance of each mid-point from the mean, 64.80. No attention need be given the sign of this distance. In the fifth column is the $\sigma$ distance of each mid-point from the mean, obtained by dividing the entry in the previous column by $\sigma$, 9.32. Next the theoretical frequency for each class is found. It is determined by multiplying the height, as given in Table XLIV, at the $\sigma$ distance by the maximum ordinate, here 160.50. Thus the tabled height of the normal curve at $3.51\sigma$ from the mean is .0021 and this times 160.50 equals .34.[2] The following column contains the difference between each actual and corresponding theoretical frequency, taken without regard to sign. To avoid the use of too small frequencies, those in the first three classes have been combined, as also have those in the last three. Therefore the first difference is the sum of the first three actual frequencies minus that of the first three theoretical frequencies. This gives $(1 + 3 + 9) - (.34 + 1.96 + 8.22) = 13. - 10.52 = 2.48$ as the first difference. The second is 4.40, that between 22. and 26.40, and so on for the others.

$\chi^2$ may be found by the formula

$$\chi^2 = \Sigma\left(\frac{\text{diff}^2}{f}\right),$$

[2] Another method of finding theoretical frequencies may be employed. It is to find the $\sigma$ distance of each class limit, rather than of the mid-point, then the normal areas beyond each of the two limits of each class, and finally to take the difference of these areas times $N$.

in which $f$ is the theoretical frequency. Thus the first entry in the next column is .59, obtained by dividing $2.48^2$ by 10.52; the next is .73, which is $4.40^2 \div 26.40$; and similarly for the others. Their sum—that is, $\chi^2$—is 8.83.

To interpret this, or any value of $\chi^2$, a table such as XXXVII is needed. It gives values of $P$, probability, corresponding to those of $\chi^2$ and numbers of classes.[3] In this case there are 10 classes, since the number remaining after combinations of small classes should be used. Interpolating in the table gives $P = .46$ for $\chi^2 = 8.83$ and 10 classes. This means that there are 46 chances out of 100, or 46 to 54, or about 1. to 1.17, that random sampling would yield a fit no better than that in this case.

Although interpretation in terms of adjectives is not very satisfactory, Culler's suggestions may be helpful. They are that if $P$ is .75 or more, the fit of the actual to the theoretical distribution is superlative; if $P$ is .50 to .74, it is excellent; if .25 to .49, good; if .10 to .24, fair; if .05 to .09, poor; if .00 to .04, unacceptable. Another suggestion is that if $P$ is less than .20 the distribution probably resembles some other type of curve more closely than it does the normal.

The last three columns of Table XXXVI illustrate further, but rarely worth-while, analysis of the situation. It is the determination of the chances that the difference between the actual and the theoretical frequency of each class is significant. The first of these three columns contains the standard error of this difference. It is found by the formula

$$\epsilon_{diff} = \sqrt{\frac{f(N-f)}{N}} \quad \text{or} \quad \sqrt{f - \frac{f^2}{N}},$$

[3] Degrees of freedom instead of classes may be used. Degrees of freedom may be defined as the number of independent observations or variates minus the number of true measures or parameters computed or estimated from them. In the case of grouped distributions such as the one above, for which the three measures $N$, $M$, and $\sigma$ are found, the number of degrees of freedom is regularly three less than the number of classes. In this special case, however, it is best taken as only one less. Justification of this may be found in Peters and Van Voorhis, also elsewhere. See Peters, Charles C., and Van Voorhis, Walter R., *Statistical Procedures and Their Mathematical Bases*. Pages 412–414. New York: McGraw-Hill Book Company, Inc., 1940.

TABLE XXXVII

VALUES OF $P$ CORRESPONDING TO GIVEN VALUES OF $\chi^2$ AND NUMBERS OF CLASSES*

| $\chi^2$ | Number of Classes | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
| 1 | .32 | .61 | .80 | .91 | .96 | .99 | .99 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| 2 | .16 | .37 | .57 | .74 | .85 | .92 | .96 | .98 | .99 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| 3 | .09 | .22 | .39 | .56 | .70 | .81 | .89 | .93 | .96 | .98 | .99 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| 4 | .05 | .14 | .26 | .41 | .55 | .68 | .78 | .86 | .91 | .95 | .97 | .98 | .99 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| 5 | .03 | .08 | .17 | .29 | .42 | .54 | .66 | .76 | .83 | .89 | .93 | .96 | .98 | .99 | .99 | 1.00 | 1.00 | 1.00 | 1.00 |
| 6 | .02 | .05 | .11 | .20 | .31 | .42 | .54 | .65 | .74 | .82 | .87 | .92 | .95 | .97 | .98 | .99 | .99 | 1.00 | 1.00 |
| 7 | .01 | .03 | .07 | .14 | .22 | .32 | .43 | .54 | .64 | .73 | .80 | .86 | .90 | .93 | .96 | .97 | .98 | .99 | .99 |
| 8 | .00 | .02 | .05 | .09 | .16 | .24 | .33 | .43 | .53 | .63 | .71 | .79 | .84 | .89 | .92 | .95 | .97 | .98 | .99 |
| 9 | .00 | .01 | .03 | .06 | .11 | .17 | .25 | .34 | .44 | .53 | .62 | .70 | .77 | .83 | .88 | .91 | .94 | .96 | .97 |
| 10 | .00 | .01 | .02 | .04 | .08 | .12 | .19 | .27 | .35 | .44 | .53 | .62 | .69 | .76 | .82 | .87 | .90 | .93 | .95 |
| 11 | .00 | .00 | .01 | .03 | .05 | .09 | .14 | .20 | .28 | .36 | .44 | .53 | .61 | .69 | .75 | .81 | .86 | .89 | .92 |
| 12 | .00 | .00 | .01 | .02 | .03 | .06 | .10 | .15 | .21 | .29 | .36 | .45 | .53 | .61 | .68 | .74 | .80 | .85 | .89 |
| 13 | .00 | .00 | .00 | .01 | .02 | .04 | .07 | .11 | .16 | .22 | .29 | .37 | .45 | .53 | .60 | .67 | .74 | .79 | .84 |
| 14 | .00 | .00 | .00 | .01 | .02 | .03 | .05 | .08 | .12 | .17 | .23 | .30 | .37 | .45 | .53 | .60 | .67 | .73 | .78 |
| 15 | .00 | .00 | .00 | .00 | .01 | .02 | .04 | .06 | .09 | .13 | .18 | .24 | .31 | .38 | .45 | .52 | .60 | .66 | .72 |
| 16 | .00 | .00 | .00 | .00 | .01 | .01 | .03 | .04 | .07 | .10 | .14 | .19 | .25 | .31 | .38 | .45 | .52 | .59 | .66 |
| 17 | .00 | .00 | .00 | .00 | .00 | .01 | .02 | .03 | .05 | .07 | .11 | .15 | .20 | .26 | .32 | .39 | .45 | .52 | .59 |
| 18 | .00 | .00 | .00 | .00 | .00 | .01 | .01 | .02 | .04 | .05 | .08 | .12 | .16 | .21 | .26 | .32 | .39 | .46 | .52 |
| 19 | .00 | .00 | .00 | .00 | .00 | .00 | .01 | .01 | .03 | .04 | .06 | .09 | .12 | .16 | .21 | .27 | .33 | .39 | .46 |
| 20 | .00 | .00 | .00 | .00 | .00 | .00 | .01 | .01 | .02 | .03 | .05 | .07 | .10 | .13 | .17 | .22 | .27 | .33 | .39 |
| 21 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .01 | .01 | .02 | .03 | .05 | .07 | .10 | .14 | .18 | .23 | .28 | .34 |
| 22 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .01 | .02 | .02 | .04 | .06 | .08 | .11 | .14 | .18 | .23 | .28 |
| 23 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .01 | .01 | .02 | .03 | .04 | .06 | .08 | .11 | .15 | .19 | .24 |
| 24 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .01 | .01 | .02 | .03 | .05 | .07 | .09 | .12 | .16 | .20 |
| 25 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .01 | .01 | .01 | .02 | .03 | .05 | .07 | .09 | .12 | .16 |
| 30 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .01 | .01 | .02 | .03 | .04 | .05 |
| 35 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .01 | .01 | .02 |
| 40 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 |

* This table might be given according to numbers of degrees of freedom rather than of classes. In this case the number of degrees of freedom should be taken as one less than the number of classes.

in which $f$ is, as above, the theoretical frequency. Thus, for the first difference, 2.48,

$$\epsilon_{diff} = \sqrt{\frac{10.52(750. - 10.52)}{750.}} = 3.22.$$

In the next column are the quotients of the differences divided by their respective standard errors, such as $2.48 \div 3.22 = .77$. By considering this as an entry in the first column of Table XLIV and finding the corresponding entry in the last column of the same table, we obtain the chances to 1 that the difference between actual and theoretical frequencies of a single class is significant. For .77 they are 3.5; for the next entry, .87, they are 4.2; and so on. For only one class, that from 65 up to 70, are the chances great enough to approach what is usually considered reasonable certainty.

The procedure just described is not limited in its application to comparisons with a normal distribution, but may be employed for those with others also. Likewise it may be used with data in such tables as those from which tetrachoric, phi, and contingency coefficients are computed. In such cases what is known as an *independence table* is set up and the obtained data compared with it. An independence table is one that contains a purely random distribution within its cells of the cases composing the class totals; in other words, it is a table showing how the given numbers of cases would be distributed if there were zero correlation between the variables or attributes concerned.

To illustrate the application of the $\chi^2$ test to such a situation, the data used for the computation of the coefficient of contingency will be employed. The entries in a number of the cells are less than 5, which is not desirable, but the procedure will be applied so as to show the method. The effect of grouping small frequencies together tends to be to increase $P$, which in this instance is found to be so high without doing that as to leave slight possibility of increase. The first part of Table XXXVIII repeats the data in the same part of Table XXXII on page 190. Part B is the corresponding independence table. Each entry therein is the product of the frequencies of the row

## TABLE XXXVIII

APPLICATION OF $\chi^2$ TEST TO CONTINGENCY TABLE BY MEANS OF INDEPENDENCE TABLE

PART A. ACTUAL DATA

| Mark | Major Subject | | | | | | |
|---|---|---|---|---|---|---|---|
| | English | For. Lang. | Ind. Arts | Math. | Science | Soc. St. | Total |
| *A* | 4 | 3 | 6 | 2 | 3 | 3 | 21 |
| *B* | 8 | 4 | 13 | 6 | 8 | 6 | 45 |
| *C* | 7 | 4 | 10 | 7 | 8 | 5 | 41 |
| *D* | 5 | 1 | 4 | 5 | 3 | 3 | 21 |
| *E* | 1 | 0 | 1 | 1 | 2 | 1 | 6 |
| Total | 25 | 12 | 34 | 21 | 24 | 18 | 134 |

PART B. INDEPENDENCE TABLE

| Mark | Major Subject | | | | | | |
|---|---|---|---|---|---|---|---|
| | English | For. Lang. | Ind. Arts | Math. | Science | Soc. St. | Total |
| *A* | 3.92 | 1.88 | 5.33 | 3.29 | 3.76 | 2.82 | 21 |
| *B* | 8.40 | 4.03 | 11.42 | 7.05 | 8.06 | 6.04 | 45 |
| *C* | 7.65 | 3.67 | 10.40 | 6.43 | 7.34 | 5.51 | 41 |
| *D* | 3.92 | 1.88 | 5.33 | 3.29 | 3.76 | 2.82 | 21 |
| *E* | 1.12 | .54 | 1.52 | .94 | 1.07 | .81 | 6 |
| Total | 25 | 12 | 34 | 21 | 24 | 18 | 134 |

PART C. COMPUTATION OF $\chi^2$

| Mark | Major Subject | | | | | | |
|---|---|---|---|---|---|---|---|
| | English | For. Lang. | Ind. Arts | Math. | Science | Soc. St. | Total |
| *A* | .00 | .67 | .08 | .51 | .15 | .01 | 1.42 |
| *B* | .02 | .00 | .22 | .16 | .00 | .00 | .40 |
| *C* | .06 | .03 | .02 | .05 | .06 | .05 | .27 |
| *D* | .30 | .41 | .33 | .89 | .15 | .01 | 2.09 |
| *E* | .01 | .54 | .18 | .00 | .81 | .04 | 1.58 |
| Total | .39 | 1.65 | .83 | 1.61 | 1.17 | .11 | 5.76 |

$\chi^2 = 5.76$ $P = 1.00-$

and column wherein it falls divided by the total number of cases. Thus for a mark of $A$ in English the independence value is $\frac{21 \times 25}{134} = 3.92$. This is the number of $A$'s in English that would result if there were no association at all between subjects and marks. Similarly all the other independence values are found. In Part C are the squares of the differences divided by the independence values. For example, for $A$'s in English the difference is $4. - 3.92 = .08$ and $.08^2 \div 3.92 = .00$ to two places. For $A$'s in foreign language, $3. - 1.88 = 1.12$, and this squared divided by 1.88 gives .67, and so on for the other cells. The sum of these entries is 5.76, which is $\chi^2$.

If the coefficient of contingency has been found, $\chi^2$ can be computed very easily. It is merely $N(T - 1)$. For these data $N$ is 134 and $T$ was found to be 1.0430; hence

$$\chi^2 = 134.(1.0430 - 1.) = 5.76,$$

as before. If $T$ is not available, but only the coefficient of contingency itself, $\chi^2$ can be found by the formula $\frac{NC^2}{1 - C^2}$. In this instance this gives

$$\chi^2 = \frac{134. \times .20^2}{1. - .20^2} = 5.58,$$

which differs from 5.76 because $C$ was computed and used to only two places. Conversely, if $\chi^2$ is known and $C$ is desired, the formula

$$\sqrt{\frac{\chi^2}{\chi^2 + N}}$$

may be employed. Thus here,

$$C = \sqrt{\frac{5.76}{5.76 + 134.}} = .20.$$

Although Table XXXVII was constructed for use with grouped distributions rather than two-way tables, it may be employed with the latter by finding the number of degrees of freedom, adding 1, and considering this as the number of classes. The number of degrees of freedom in a two-way table is regularly the product of $n_1 - 1$ and $n_2 - 1$, $n_1$ being the number of

classes in one direction and $n_2$ that in the other. Here this gives $(6 - 1)(5 - 1) = 20$ degrees of freedom, corresponding to 21 classes. Table XXXVII ends with 20 classes, but it is evident that for $\chi^2 = 5.76$ and 21 classes $P$ is nearer 1.00 than .99. In other words, there are practically 100 chances in 100 that the differences of actual from theoretical data are mere random variations and do not signify any real tendency toward relationship between marks and school subjects.

As another example, the data employed for tetrachoric correlation on page 192 may also be subjected to the $\chi^2$ test. They, the independence table for them, and the determination of $\chi^2$ appear in Table XXXIX. Since the steps are just the

TABLE XXXIX

APPLICATION OF $\chi^2$ TEST TO TETRACHORIC TABLE

PART A. BY MEANS OF INDEPENDENCE TABLE

| ACTUAL DATA | | | | INDEPENDENCE TABLE | | | | COMPUTATION OF $\chi^2$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Element No. 2 | Element No. 1 | | | Element No. 2 | Element No. 1 | | | Element No. 2 | Element No. 1 | | |
| | *R* | *W* | *T* | | *R* | *W* | *T* | | *R* | *W* | *T* |
| *W* | 67. | 63. | 130. | *W* | 77.5 | 52.5 | 130 | *W* | 1.42 | 2.10 | 3.52 |
| *R* | 82. | 38. | 120. | *R* | 71.5 | 48.5 | 120. | *R* | 1.54 | 2.27 | 3.81 |
| *T* | 149. | 101. | 250. | *T* | 149.0 | 101.0 | 250. | *T* | 2.96 | 4.37 | 7.33 |
| | | | | | | | | $\chi^2 = 7.33$ | | $P = .01$ | |

PART B. BY MEANS OF FRACTIONAL ENTRIES

| Element No. 2 | Element No. 1 | | |
|---|---|---|---|
| | *R* | *W* | *T* |
| *W* | .268 | .252 | .520 |
| *R* | .328 | .152 | .480 |
| *T* | .596 | .404 | 1.000 |

$$\chi^2 = \frac{250.(.268 \times .152 - .252 \times .328)^2}{.596 \times .404 \times .520 \times .480} = 7.31$$

$P = .01$

same as in Table XXXVIII, they will not be explained again. For this, as all tetrachoric tables, there is only one degree of

freedom, which corresponds to 2 classes in Table XXXVII. Therefore for $\chi^2 = 7.33$, $P$ is about .01; so there is about 1 chance in 100 that the differences are due to variations of sampling rather than to real relationship.

For such a table $\chi^2$ may also be computed by employing frequencies expressed as fractions of the total number of cases. Using the same symbols as for the $\phi$ coefficient, the formula for this method is

$$\chi^2 = \frac{N(\alpha\delta - \beta\gamma)^2}{p_1q_1p_2q_2}.$$

For the data in Table XXXIX this gives

$$\frac{250.(.268 \times .152 - .252 \times .328)^2}{.596 \times .404 \times .520 \times .480} = 7.31,$$

which is slightly different from 7.33 because of dropping decimals.

A simpler but less exact measure of goodness of fit is the per cent of actual and theoretical frequencies that agree. One way of finding this is to compare the two for each class or cell, as the case may be, find the sum of the smaller of each such pair for the whole table, and compute its per cent of $N$. Another is to subtract half the sum of all the differences, without regard to sign, from $N$ and find what per cent the remainder is of $N$. For the data in Table XXXVI the first gives

$$10.52 + 22. + 54. + 113.70 + 153.82 + 155.38 + 105. + 63. + 28.55 + 11.86 = 717.83,$$

which divided by 750 yields 96− per cent. The second gives

$$750. - \frac{63.86}{2} = 718.07$$

and this divided by 750 likewise yields 96− per cent. These methods may also be employed for the data in Tables XXXVIII and XXXIX.

As was stated on p. 182, $\chi^2$ may be used to estimate the reliability of differences between ratios and coefficients of correlation. For this purpose

$$\chi^2 = (N - n)\frac{\eta^2 - r^2}{1 - \eta^2}.$$

For the data in Table XXIX, for which

$$\eta_{xy} = .902,\ \eta_{yx} = .914,\ r = .892,\ N = 50, \text{ and } n = 12,$$

$$\chi^2 \text{ for } \eta_{xy} = (50 - 12)\frac{.902^2 - .892^2}{1. - .902^2} = 3.60$$

and

$$\text{for } \eta_{yx} = (50 - 12)\frac{.914^2 - .892^2}{1. - .914^2} = 9.17.$$

Employing these values in Table XXXVII, with the number of classes taken as $n - 1$ or 11, yields $P = .97$ and .51, respectively. There are, therefore, about 97 chances out of 100 that the excess of curvilinear over rectilinear relationship is due to chance in the first case, and 51 out of 100 that it is in the second.

## Determining difficulties of test elements

For purposes of weighting and other reasons those who construct and use tests may desire to determine the relative difficulties of test elements. The generally accepted method, one that has been employed for many standard tests, is based on the assumption that, if the area under the normal curve is divided into two portions—one corresponding to the fraction of individuals passing an element and the other to that of those failing it—the point of division is an index of difficulty. This point is expressed in terms of some measure of variability, usually either the standard or the median deviation. This assumption has not been proved, but in practice appears to give reasonably valid results.

The details of applying the procedure just suggested vary somewhat. Either fraction passing or fraction failing, or sometimes one and sometimes the other, may be employed. Perhaps the easiest way is to use whichever is the smaller, locate it in the fourth column of the table of the normal curve in the Appendix, find the corresponding entry in the first column, and prefix a plus sign to it if the fraction passing was used, a minus sign if the fraction failing was taken. This should be done because, if less than half pass an element, it is more difficult than average; whereas, if less than half fail it. it is less difficult.

To illustrate this, let us suppose that 85 per cent of those taking a test pass and 15 per cent fail a certain element. The $\sigma$ distance corresponding to 15 per cent of area beyond an ordinate at that distance is about 1.04; hence the element is $1.04\sigma$ less difficult than average, so receives a difficulty rating of $-1.04\sigma$. For one passed by 30 per cent and failed by 70 per cent we find that about $.52\sigma$ corresponds to an area of .30; hence rate it $+.52\sigma$. If an element is passed or failed by everyone, it is rated at $-\infty$ or $+\infty$.

In order to eliminate minus signs, some point far enough below average so that no element is likely to be easier than one rated there is often chosen and difficulty values expressed as distances above that. Thus, for the well-known and much used *T*-scale and *T*-scores,[4] based on a normal distribution of twelve-year-olds, a point $5\sigma$ below the mean was chosen. Accordingly a difficulty value of $-1.04\sigma$ becomes $5. - 1.04 = 3.96$; one of $+.52\sigma$ becomes $5. + .52 = 5.52$; and so on. Also a multiplier, 10, was used and fractions dropped, thus giving a scale ranging from 0, with 50 as average, up to 100. On this scale a difficulty value of 3.96 as given above becomes 40; one of 5.52 becomes 55; and so on.

## Equally noticed differences

The theorem that equally noticed differences are equal, with certain concomitant additions, is sometimes applied to determine the quality or merit of specimens, such as those used in handwriting, drawing, composition, and other quality scales. The validity of its application depends upon several assumptions, the chief of which are that the distributions of ratings of the various specimens are normal in shape and that areas under the normal curve and base-line distances validly represent differences in raters' judgments and in quality, respectively. In general, the results from this method correlate so highly with those obtained by using average ratings on a quantitative scale that which is employed is a matter of convenience rather than of validity of results.

---

[4] McCall suggested this procedure and named it in honor of Thorndike and Terman.

To illustrate this method, let us assume that we have had a group of raters or judges compare specimen A with B, B with C, and C with D, and that 61 per cent of them have rated B as better than A, 75 per cent have rated C as better than B, and 94 per cent have rated D as better than C.[5] The following procedure is merely to subtract 50 per cent from each per cent of better ratings and find the deviation distance corresponding to this area from the mean. Generally the median deviation is used, although any other can be used. For A and B, we have 61 per cent − 50 per cent = 11 per cent, and from Table XLV in the Appendix we find that the median deviation distance that includes this fraction of the area between an ordinate at its distance from the mean and the mean is approximately .41; therefore specimen B is $.41MdD$ better than specimen A. By similar procedure, C is $1.00MdD$ better than B, and D is $2.30MdD$ better than C. Likewise, C is $.41MdD + 1.00MdD = 1.41MdD$ better than A; D is $.41MdD + 1.00MdD + 2.30MdD = 3.71MdD$ better than A; and D is $1.00MdD + 2.30MdD = 3.30MdD$ better than B.

If it is desired, each possible pair may be compared and the results obtained from the various direct and indirect comparisons may be averaged. Thus, if 99 per cent of the raters placed D above A, the corresponding difference in merit would be $3.45MdD$ rather than the $3.71MdD$ found indirectly through B and C. An average of the two, 3.58, is probably better than either. If many specimens are involved, the number of possible indirect comparisons becomes so great as to render the task of making and averaging them so laborious that it is rarely carried out in full.

If the per cent of judges rating one sample as better than other is 100, the difference cannot be determined more specifically than as approaching infinity. If it is 50, the other 50 per cent evidently consider it worse; hence the two specimens are of equal merit. If it is less than 50, one may either reverse the direction of comparison or take the difference between it and

---

[5] In this method raters must give no tied or equal ratings, but must always rate one of the pair of specimens being compared above the other.

50 and interpret the corresponding *MdD* distance as the amount the sample in question is below the other.

In connection with this procedure we often wish to know how many judges are necessary to result in a given desired degree of reliability. This depends on both their number and the per cent that rate one sample above the other. If the reliability of ratings from a given number of judges is known, that for some other number is given approximately by the Spearman-Brown formula, found on page 151. More exactly, however, the number of judges necessary for a given degree of reliability—that is, for a given probability that a specimen voted better than another would likewise be rated better by another similar group of judges—is the next larger integer than the value of the formula

$$\frac{x^2pq}{(p - .5)^2},$$

in which $x$ is the number of standard errors corresponding to the degree of reliability, $p$ is the fraction of judges rating one specimen better than the other, and $q$ is $1 - p$, or the fraction of judges rating it worse. Table XL gives the numbers of judges, as determined by this formula, necessary to insure the

TABLE XL

NUMBERS OF JUDGES NECESSARY TO INSURE VARIOUS DEGREES OF RELIABILITY OF RATINGS OF SPECIMENS

| Standard Errors | Chances to 1 That Specimen Rated Better Is Better | Per Cent of Judges Rating One Specimen Better than Other | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | 55 | 60 | 65 | 70 | 75 | 80 | 85 | 90 | 95 |
| 3.0 | 739 | 892 | 217 | 92 | 48 | 28 | 17 | 10 | 6 | 3 |
| 2.5 | 161 | 619 | 151 | 64 | 33 | 19 | 12 | 7 | 4 | 2 |
| 2.0 | 43 | 397 | 97 | 41 | 22 | 13 | 8 | 5 | 3 | 1 |
| 1.5 | 14 | 223 | 55 | 23 | 12 | 7 | 5 | 3 | 2 | 1 |
| 1.0 | 5.3 | 100 | 25 | 11 | 6 | 4 | 2 | 2 | 1 | 1 |
| .5 | 2.2 | 25 | 7 | 3 | 2 | 1 | 1 | 1 | 1 | 1 |

chances, corresponding to certain values of the standard error and per cents of judges rating one specimen above another, that the specimen so rated really is the better of the two. It makes apparent that the number of judges needed decreases quite rapidly as the per cent rating one specimen above another increases.

## Equivalent or comparable measures

Although the topic of equivalent or comparable measures has been touched previously, it seems well to present a short summary treatment here. Very frequently educational workers find it desirable to compare or combine scores derived from various tests and other sources with different scoring systems and units. To do so more than very superficially necessitates that they be transmuted to a common system, that is, be expressed on equivalent scales with equivalent units.

If the distributions of scores involved are nearly normal, or only moderately skew, the standard or $z$-score is generally the most valid comparable score. As already given, it is

$$z = \frac{\text{Score} - M}{\sigma},$$

that is, it expresses scores in terms of their $\sigma$ distances from the mean. It assumes that the mean score of one distribution may be considered to correspond to that of the other, and similarly for their standard deviations. Similar scores based upon some other appropriate combination of a measure of central tendency and one of variability could be employed, but rarely are.

Another means of rendering scores comparable is the use of the *equivalent score equations*, which are simply the regression equations with $r = 1.00$. By their use scores in other distributions may be changed so as to be expressed in the same units and on the same scale as those in one taken as standard; or an ideal distribution with convenient values, such as $M = 50$ and $\sigma = 10$, may be set up and used. Results from the application of this and the $z$-score method to the same original scores correlate perfectly. In general it is merely a matter of convenience which is used. Since the $z$-score method is much

more commonly employed and is probably easier, it appears preferable.

If distributions approximate rectangular form, percentile ranks, or just ordinary ranks if the numbers of cases involved are the same, are the best. They are often employed regardless of the shape of distributions, but so doing is liable to lead to erroneous conclusions. If a distribution approximates normality, *U*-shape, or any of various other types, the differences between percentile ranks or points apparently equidistant are not equal. For example, in such cases the difference between $P_{90}$ and $P_{80}$ is not the same as that between $P_{80}$ and $P_{70}$ or as that between $P_{60}$ and $P_{50}$. In a rectangular distribution such differences are equal.

A third type of comparable measure that is more often employed elsewhere than in educational work is the *ratio score*. It is valid only if three conditions are fulfilled: the zero points of the series of scores being compared must represent the same amounts of whatever they measure; some other point, usually the mean, of one series must correspond to the same point of the other; and the laws of development, that is, the shapes of the distributions of scores, of the things measured must be the same. The third condition is often not true, or at least not known to be true, and the other two are far from universally so; hence the validity of ratio scores is limited.

The formula for a ratio score is simply $\frac{X}{B}$, in which $X$ represents the score and $B$ the base. Often, as suggested above, this becomes $\frac{X}{M}$. To illustrate the use of ratio scores, we may assume that a distribution of pupils' scores upon an achievement test and that of their mental ages meet the three conditions stated above, and that the means are 65 and 11–8, respectively. If pupil A makes 56 on the test and has a *MA* of 11–4, his ratios are $\frac{56}{65} = .86$ and $\frac{11-4}{11-8} = .97$.[6] A comparison of these ratios

[6] The easiest way to handle this ratio is to reduce the *MA*'s to months. Thus 11–4 = 136 months and 11–8 = 140 months, and $\frac{136}{140} = .97$.

indicates that he stands distinctly lower on the test than in $MA$. If pupil B has a score of 83 and a $MA$ of 15−1, his ratios are $\frac{83}{65} = 1.28$ and $\frac{15-1}{11-8} = 1.29$, which indicate that his test score and $MA$ are relatively very nearly the same.

## EXERCISES AND PROBLEMS

**1.** Find the number of cases falling between each of the following pairs of scores, assuming that each distribution is normal and has a mean, standard, or median deviation, and number of cases as given:

(*a*) Between 40 and 50, if $M = 34.5$, $\sigma = 7.8$, and $N = 90$.
(*b*) Between 60 and 80, if $M = 63.6$, $\sigma = 12.2$, and $N = 125$.
(*c*) Between 5 and 32, if $M = 49.$, $MdD = 8.4$, and $N = 420$.

**2.** Determine the per cents of pupils who should receive each mark if they are normally distributed according to the following specifications:

(*a*) Six marks; curve cut off at $\pm 2.5\sigma$.
(*b*) Five marks; curve cut off at $\pm 4.0\ MdD$.
(*c*) Seven marks; curve cut off at $\pm 3.0\sigma$.

**3.** Assign marks according to the $z$-score method to the scores given in each part of Exercise 6 on page 70. Assume the following bases for the several parts:

(*a*) Five marks; curve cut off at $\pm 2.5\sigma$.
(*b*) Six marks; curve cut off at $\pm 3.0\sigma$.
(*c*) Four marks; curve cut off at $\pm 3.0\sigma$.

**4.** Change each of the following sets of data into a normal distribution, with $\sigma$ values for each class:

(*a*) *High*, 5; *good*, 17; *fair*, 23; *poor*, 8.
(*b*) *Superior*, 16; *good*, 38; *average*, 64; *poor*, 27; *failing*, 6.
(*c*) *A*, 14; *B*, 18; *C*, 24; *D*, 15; *E*, 9; *F*, 2.

**5.** Find the probability that each of the distributions in Exercise 1 on page 69 differs from normal.

**6.** Find the probability that each of the sets of data in Exercises 4 on page 196 and 6 on page 197 differ from an independence table.

**7.** Determine the difficulty of test elements answered incorrectly by each of the following per cents of pupils: (*a*) 3; (*b*) 17; (*c*) 81; (*d*) 46; (*e*) 99; (*f*) 65.

**8.** If three supervisors assign ratings to a group of 50 teachers as shown below, find the average $\sigma$ rating of each of the two teachers whose ratings are given.

| Supervisor | Ratings | | | | |
|---|---|---|---|---|---|
| | *A* | *B* | *C* | *D* | *E* |
| Brown | 3 | 12 | 18 | 10 | 7 |
| Jones | 4 | 7 | 28 | 8 | 3 |
| Smith | 10 | 15 | 15 | 8 | 2 |

Teacher No. 1: *B* by Brown, *C* by Jones, *A* by Smith.
Teacher No. 2: *D* by Brown, *C* by Jones, *C* by Smith.

**9.** Find the difference in quality of pairs of specimens, one of which is rated better than the other by each of the following per cents of judges: (*a*) 68; (*b*) 82; (*c*) 36; (*d*) 97.

**10.** The following are thc scores made by a class of 40 pupils on a series of tests, one in each of several subjects:

| *Reading* | *Arithmetic* | *Language* | *Writing* | *Spelling* |
|---|---|---|---|---|
| 50–2 | 24–1 | 70– 1 | 60– 2 | 45– 3 |
| 45–3 | 22–0 | 60– 4 | 50– 6 | 40–14 |
| 40–4 | 20–3 | 50–12 | 40–21 | 35–11 |
| 35–6 | 18–4 | 40–16 | 30– 8 | 30– 6 |
| 30–8 | 16–3 | 30– 4 | 20– 3 | 25– 5 |
| 25–7 | 14–9 | 20– 2 | | 20– 1 |
| 20–6 | 12–6 | 10– 1 | | |
| 15–2 | 10–5 | | | |
| 10–2 | 8–5 | | | |
| | 6–2 | | | |
| | 4–0 | | | |
| | 2–0 | | | |
| | 0–2 | | | |

Find the average standing, by both *z*-scores and percentile scores, of pupils whose scores on the five tests are as follows:

| *Pupil* | *Reading* | *Arithmetic* | *Language* | *Writing* | *Spelling* |
|---|---|---|---|---|---|
| A | 36 | 21 | 51 | 44 | 40 |
| B | 23 | 9 | 26 | 33 | 29 |
| C | 52 | 19 | 61 | 53 | 48 |
| D | 13 | 7 | 25 | 22 | 30 |
| E | 33 | 17 | 44 | 42 | 36 |

**11.** Find the chances that each one of the pair rated better in Exercise 9 really is better if each of the following numbers of judges participated in rating: (*a*) 10; (*b*) 75; (*c*) 24; (*d*) 3.

**12.** Find the *z*-score, the percentile score, and the ratio score, using the mean as the base, of each of the following pairs of scores. The first one of each pair is a score in the column headed "Fast" and the second in that headed "Aver." in Part (*a*) of Exercise 3 on page 80. From the comparable scores found, determine which of each pair is relatively higher with regard to its own group.

(*a*) 132, 115; (*b*) 124, 106; (*c*) 112, 92.

**13.** Do the same with these scores from Part (*b*) of Exercise 3 on page 80.

(*a*) 95, 100; (*b*) 79, 86; (*c*) 103, 109.

## REFERENCES

Fisher, R. A., *Statistical Methods for Research Workers.* Eighth Edition. Chap. IV; pages 41–59. Edinburgh: Oliver and Boyd, 1941.

Garrett, Henry E., *Statistics in Psychology and Education.* Second Edition. Pages 119–126, 132–173, 177–191, 577–587. New York: Longmans, Green & Company, Inc., 1937.

Guilford, J. P., *Fundamental Statistics in Psychology and Education.* Chap. VII; pages 81–92, 167–173. New York: McGraw-Hill Book Company, Inc., 1942.

Kenney, John F., *Mathematics of Statistics.* Part II, pages 164–171. New York: D. Van Nostrand Company, Inc., 1939.

Lindquist, E. F., *A First Course in Statistics.* Revised Edition. Chap. IX; pages 85–101. Boston: Houghton Mifflin Company, 1942.

Lindquist, E. F., *Statistical Analysis in Educational Research.* Chap. II. Boston: Houghton Mifflin Company, 1940.

Mode, Elmer B., *The Elements of Statistics.* Pages 356–361. New York: Prentice-Hall, Inc., 1941.

Monroe, Walter S., Editor, *Encyclopedia of Educational Research.* Pages 341–344. New York: The Macmillan Company, 1941.

Monroe, Walter S., and Engelhart, Max D., *The Scientific Study of Educational Problems.* Pages 80–85. New York: The Macmillan Company, 1926.

Peters, Charles C., and Van Voorhis, Walter R., *Statistical Procedures and Their Mathematical Bases.* Pages 304–311, 404–419. New York: McGraw-Hill Book Company, Inc., 1940.

Walker, Helen M., *Elementary Statistical Methods.* Pages 176–198. New York: Henry Holt & Company, Inc., 1943.

## CHAPTER XIV

# Sampling

### Nature and use of sampling

All operations with data of any sort may be classed as belonging to one of two general types: either the data involved include all of the sort in which the workers are interested; or they are a sample of a larger body of data which it is impossible or impracticable to include in its entirety. Especially in research, as contrasted with routine dealing with facts, is the latter often true. In many cases the same data may be employed in either way, or in both. Thus, if the heights of a class of forty children are the data used, they may be utilized either to determine desired facts about just that group of children or to furnish evidence as to probable facts for a larger group of which the group of forty is supposed to be representative. When used in the former way, they are frequently referred to as *descriptive statistics;* when in the latter, the drawing of conclusions about the larger group is often called *statistical inference.* In the latter case, and all others in which data from a sample are considered indicative of facts concerning the total group it represents, there are likely to be errors due to the fact that the sample is not perfectly representative of the whole group, or *population* or *universe,* as it is often called.

The terms *statistic* and *parameter* are so frequently used in this connection that readers should know them. Each term has other meanings, but as employed here a *statistic* is a measure, such as a mean, a median, a standard deviation, a coefficient of correlation, and so on, computed from a sample; whereas a *parameter* is a similar measure of a whole population. In other words, a statistic is an actually secured measure subject

to error when considered as a measure of a universe; whereas a parameter is a true measure thereof.[1]

The whole theory of sampling errors has to do with variable rather than systematic errors. In other words, it is based upon the supposition that samples are chosen by some random or purely chance method. A random method is one according to which every case in the population sampled has an equal chance of being chosen. It eliminates all intentional bias and generally most unintentional bias also. The theory of sampling attempts to describe the extent to which such a random sample is, or is not, truly representative of the population from which it was chosen. This chapter deals with the estimation of the size and frequency of *errors of sampling* or, in other words, with the reliability of measures secured from samples when taken as indicative of similar measures of universes. Untrained workers often commit the mistake of computing errors of sampling when data from the total population have been used. In such instances, no sampling has occurred and consequently no such errors can be present.

## Selecting samples

In selecting samples, especially small ones, workers should observe certain precautions which tend to prevent bias rather than rely entirely upon chance. These precautions are of two chief kinds: those having to do with the way in which tentative samples are chosen, and those concerned with checking their representativeness before employing them. The basis of selecting samples should be one that has no correlation with the characteristics to be measured. In judging this point workers should be critical and should look beneath the surface.

If the population to be sampled contains a number of groups or subdivisions of which the sizes are known, it is better to draw a random sample proportional to the size of the group from each, and then combine these, than to select a single random sample from the whole population. This method is sometimes

[1] To indicate that a measure applies to a universe rather than to a sample, a curl ($\sim$) may be written above it. For example, $\tilde{\sigma}$ stands for the standard deviation of a universe, not of a sample. Other methods of distinguishing between them are sometimes employed, but this appears to be the most satisfactory.

called *sampling by subdivision* or *by stratification.* For example, if the population is known to contain 54 per cent girls and 46 per cent boys and a 10 per cent sample is to be taken, it is better to choose 10 per cent of each sex group by some random method than merely to select 10 per cent of the total number by the same method. If this procedure has not been followed at first, after a tentative sample has been picked from the whole population it should be checked to determine if it represents the known subdivisions equally well.

To illustrate sampling further, let us suppose that a superintendent of schools of a populous county wishes to measure the general achievement of pupils completing the elementary school, but cannot well expend the money and effort necessary to test more than one-fifth of such pupils. A good method of selecting those to be tested would be to write the names of all graduating pupils upon similar slips and, after mixing them thoroughly, to select one-fifth by a purely chance procedure. A better one would be to arrange the names of all alphabetically and take every fifth one. Even better would be to apply the same method by schools or districts rather than to the county as a whole. To call for either pupils or teachers to volunteer, to take the one-fifth seated nearest the front, to test all in a few schools or systems containing the desired fraction, and to ask principals to designate pupils, are examples of unsatisfactory procedures likely to produce biased samples. Intermediate between good and bad would be to select at random one-fifth of the schools and to test all graduating pupils found in them.

An excellent sampling procedure, but one little used in educational work, is to employ a table of random numbers. A good discussion and illustration of this procedure, with such a table, is given by Lindquist.[2] A more complete table of this sort may be found among those of Fisher and Yates.[3]

It is rarely possible to determine whether or not a sample is biased merely by examining the sample itself. If comparison

---

[2] Lindquist, E. F., *Statistical Analysis in Educational Research.* Pages 25–29, 262–264. Boston: Houghton Mifflin Company, 1940.

[3] Fisher, R. A., and Yates, F., *Statistical Tables for Biological, Agricultural and Medical Research.* Pages 82–87. London: Oliver and Boyd, 1938.

with the population, as suggested above, is impossible or impracticable, the best procedure is to choose, in the same manner, but independently, several samples of equal size and then compare them. If they agree in essential characteristics, with only insignificant differences, any one, or preferably all combined, is probably satisfactory. They may be biased in the same direction, but this is unlikely. If they do not agree, either the size of the samples should be increased to a point at which they do, or else some other method of random selection should be employed.

A point often neglected in the selection of samples and the determination of the reliability of measures derived from them is that the individuals composing a random sample should not be members of a relatively few integral groups, such as classes. If they are, the sample is likely to be much less representative of the universe than if they are not. For example, a sample of 20 classes of twenty-five pupils each is much less likely to be representative of 10,000 high-school freshmen in a large city system than is one of 500 chosen by taking every twentieth one alphabetically. This is true even if the 20 classes are a thoroughly random selection of all such classes. It is almost certain that the pupils in each of the 20 classes will have certain common characteristics that affect the sample. Such a sample will tend to be intermediate in representativeness between one of 500 pupils properly chosen and one of twenty pupils, also properly chosen.

## Sampling theories

The classical or traditional method of estimating errors of sampling, often called the *large sample theory,* is based upon the assumption that such errors approximate normality in their distribution, with a mean at 0. Since most of the samples employed in educational work are of such size, greater than 30 or 40, as technically to be called *large* rather than *small* and since the distributions of errors for such samples for most common measures usually do not vary greatly from normal, this theory yields more or less satisfactory results in many cases. On the other hand, we do use some small samples, and even for large

ones the distributions of errors for some measures are not normal; hence better means of estimating sampling errors are often needed. The *small sample theory* has attempted to supply this need and has furnished formulas and interpretations based thereon that are, in some instances, distinctly superior to those previously employed. In general, the small sample theory is applicable to large samples, but when so used ordinarily yields conclusions very similar to those given by the large sample theory.

## Measures of error and their interpretation

Except in instances so rare as to be negligible, it is impossible to determine the sampling error in any single specific measure. The best that can be done is to determine the probable distribution of errors in all similar measures and to estimate the chances that the error in any one of them is of a given magnitude. For example, if the mean score of forty pupils is taken as representing the mean of a larger group from which the forty were selected, the error of sampling in that particular mean cannot be found. If it could, the mean would be corrected and the true mean found and used. Instead, the probable distribution of errors in means of all possible similarly chosen samples of forty pupils each can be estimated, also the chances that the error in the particular mean in question is more or less than a given size, or between certain limits.

Since distributions of sampling errors can be estimated, the conventional method of describing them is the same as for any distribution, by stating a measure of central tendency and one of variability. Moreover, since positive and negative errors of sampling generally balance each other, the mean may be taken as zero. Therefore stating a measure of variability alone is usually sufficient to describe a distribution of errors. The measure of variability most often employed for this purpose is the standard deviation, but the median deviation and, rarely, the mean deviation are also used. In this connection they are called the standard error, symbolized by $\sigma$ or $\epsilon$ (epsilon)[4]; the

[4] $\sigma$ rather than $\epsilon$ is the common symbol for standard error, but in order to avoid confusion with standard deviation the writer prefers, and will use, $\epsilon$.

probable error, $(PE)$; and the mean error, $(ME)$. These measures of error are interpreted just as are the corresponding measures of variability, as explained in Chapter VI. In other words, if, as usual, the distribution of errors is assumed to be normal, the tables in the Appendix may be applied to them.

In any such application and consequent interpretation, workers should remember that the chances found are those that a statistic or obtained measure is within the corresponding distance of a parameter or true measure, not vice versa. In many instances a fundamental difference in interpretation is involved.

## Classical formulas for errors of sampling

The classical or large sample theory formulas for the errors of sampling in a number of the measures presented in this volume are given in Table XLI. For some measures no such formulas have been derived. The first column of the table lists the measures or statistics for which error formulas are given; the second contains the formula for the standard error of each; and the third, that for the probable error. Just as $MdD = .6745\sigma$, so $PE = .6745\epsilon$, but the latter's value has been computed and is given in most cases. Many of the formulas are simplified approximations with such terms eliminated as largely increase their difficulty but slightly affect their values. In a few instances they cannot be simplified so as to be easily usable without too great sacrifice of accuracy; hence they are given in forms rather laborious to employ.

Although the error formulas in Table XLI are those conventionally given in connection with large samples, they are subject to a number of limitations even when so used. Some of these can be entirely or largely removed by appropriate modifications or corrections; others cannot. One of the most important is that the measures of variability, such as $\sigma$, $Q$, and so forth, which so often appear in the formulas should be parameters or measures of a universe rather than statistics or measures of a sample only. In most instances, however, only the latter are available; hence insofar as they are inaccurate estimates of the parameters, the results yielded by the given formulas are in error. In some instances corrected statistics

TABLE XLI

STANDARD AND PROBABLE ERRORS OF CERTAIN MEASURES ACCORDING TO CONVENTIONAL LARGE SAMPLE THEORY

| Measure | Standard Error ($\epsilon$) | Probable Error ($PE$) |
|---|---|---|
| Mean ($M$) | $\frac{\sigma}{\sqrt{N}}$ | $.6745\frac{\sigma}{\sqrt{N}}$ |
| Any percentile ($P_p$) | $2.5066\,\frac{\sigma\sqrt{pq}}{h_p\sqrt{N}}$ or $3.7162\,\frac{Q\sqrt{pq}}{h_p\sqrt{N}}$ | $1.6907\frac{\sigma\sqrt{pq}}{h_p\sqrt{N}}$ or $2.5066\,\frac{Q\sqrt{pq}}{h_p\sqrt{N}}$ |
| Median ($Md$) or Mid-score | $1.2533\,\frac{\sigma}{\sqrt{N}}$ or $1.8581\,\frac{Q}{\sqrt{N}}$ | $.8453\frac{\sigma}{\sqrt{N}}$ or $1.2533\frac{Q}{\sqrt{N}}$ |
| First or third quartile ($Q_1$ or $Q_3$) | $1.3626\,\frac{\sigma}{\sqrt{N}}$ or $2.0203\,\frac{Q}{\sqrt{N}}$ | $.9191\frac{\sigma}{\sqrt{N}}$ or $1.3626\frac{Q}{\sqrt{N}}$ |
| Tenth or ninetieth percentile ($P_{10}$ or $P_{90}$) | $1.7094\,\frac{\sigma}{\sqrt{N}}$ or $2.5344\,\frac{Q}{\sqrt{N}}$ | $1.1530\frac{\sigma}{\sqrt{N}}$ or $1.7094\,\frac{Q}{\sqrt{N}}$ |
| Any proportion ($p$) | $\sqrt{\frac{pq}{N}}$ or $\sqrt{\frac{p(1-p)}{N}}$ | $.6745\sqrt{\frac{pq}{N}}$ or $.6745\,\sqrt{\frac{p(1-p)}{N}}$ |
| Quartile deviation ($Q$) | $.7867\frac{\sigma}{\sqrt{N}}$ or $1.1664\frac{Q}{\sqrt{N}}$ | $.5306\frac{\sigma}{\sqrt{N}}$ or $.7867\,\frac{Q}{\sqrt{N}}$ |
| 10–90 percentile range ($D_{10-90}$) | $2.2792\frac{\sigma}{\sqrt{N}}$ or $.8892\,\frac{D_{10-90}}{\sqrt{N}}$ | $1.5373\frac{\sigma}{\sqrt{N}}$ or $.5998\,\frac{D_{10-90}}{\sqrt{N}}$ |

TABLE XLI—*Continued*

STANDARD AND PROBABLE ERRORS OF CERTAIN MEASURES ACCORDING TO CONVENTIONAL LARGE SAMPLE THEORY

| Measure | Standard Error ($\epsilon$) | Probable Error ($PE$) |
| --- | --- | --- |
| Mean deviation ($MD$) | $.6028\frac{\sigma}{\sqrt{N}}$ or $.7555\frac{MD}{\sqrt{N}}$ | $.4066\frac{\sigma}{\sqrt{N}}$ or $.5096\frac{MD}{\sqrt{N}}$ |
| Standard deviation ($\sigma$) | $.7071\frac{\sigma}{\sqrt{N}}$ or $\frac{\sigma}{\sqrt{2N}}$ | $.4769\frac{\sigma}{\sqrt{N}}$ |
| Median deviation ($MdD$) | $.4769\frac{\sigma}{\sqrt{N}}$ or $\frac{MdD}{\sqrt{2N}}$ | $.3217\frac{\sigma}{\sqrt{N}}$ or $.4769\frac{MdD}{\sqrt{N}}$ |
| Coefficient of variability ($V$) | $.7071\frac{V}{\sqrt{N}}$ or $\frac{V}{\sqrt{2N}}$ | $.4769\frac{V}{\sqrt{N}}$ |
| Skewness ($Sk$) based on $M$, $Md$, and $\sigma$ | $\frac{1.2247}{\sqrt{N}}$ or $\sqrt{\frac{3}{2N}}$ | $\frac{.8260}{\sqrt{N}}$ |
| Kurtosis ($Ku$) based on $Q$ and $D_{10-90}$ | $\frac{.2778}{\sqrt{N}}$ | $\frac{.1874}{\sqrt{N}}$ |
| Coefficient of correlation ($r$) | $\frac{1-r^2}{\sqrt{N}}$ | $.6745\frac{1-r^2}{\sqrt{N}}$ |

TABLE XLI—*Continued*

STANDARD AND PROBABLE ERRORS OF CERTAIN MEASURES ACCORDING TO CONVENTIONAL LARGE SAMPLE THEORY

| Measure | Standard Error ($\epsilon$) | Probable Error ($PE$) |
|---|---|---|
| Coefficient of regression ($b_{12}$) | $\frac{\sigma_1}{\sigma_2}\sqrt{\frac{1-r_{12}^2}{N}}$ or $\frac{\sigma_1 k_{12}}{\sigma_2\sqrt{N}}$ | $.6745\frac{\sigma_1}{\sigma_2}\sqrt{\frac{1-r^2_{12}}{N}}$ or $.6745\,\frac{\sigma_1 k_{12}}{\sigma_2\sqrt{N}}$ |
| Coefficient of reliability from Spearman-Brown formula ($r_n$) | $\frac{n(1-r^2)}{\sqrt{N}[1+r(n-1)]^2}$ | $.6745\frac{n(1-r^2)}{\sqrt{N}[1+r(n-1)]^2}$ |
| Change in length from Spearman-Brown formula ($n$) | $\frac{n(1+r)}{r\sqrt{N}}$ | $.6745\frac{n(1+r)}{r\sqrt{N}}$ |
| * Coefficient of correlation corrected for attenuation $\left(r_{corr} = \frac{r_{12}}{\sqrt{r_{1I}r_{2II}}}\right)$ | $\frac{r_{corr}}{\sqrt{N}}\sqrt{r^2_{corr} + \frac{1}{r^2_{12}} + \left(\frac{1}{4r^2_{1I}} - \frac{r^2_{1I}}{4} + r_{1I} - \frac{1}{r_{1I}} - 1\right) + \left(\frac{1}{4r^2_{2II}} - \frac{r^2_{2II}}{4} + r_{2II} - \frac{1}{r_{2II}} - 1\right)}$ | |
| * Same $\left(r_{corr} = \frac{\sqrt[4]{r_{12}r_{1I1}r_{12}r_{1II}}}{\sqrt{r_{1I}r_{2II}}}\right)$ | $\frac{r_{corr}}{2\sqrt{N}}\sqrt{4r^2_{corr} + \frac{1}{r^2_{corr}} + \frac{1 + r_{1I} + r_{2II}}{\sqrt{r_{12}r_{1I1}r_{12}r_{1II}}} + \frac{1}{r^2_{1I}} + \frac{1}{r^2_{2II}} - \frac{4}{r_{1I}} - \frac{4}{r_{2II}} - 2}$ | |

* In these cases the standard error alone is given. It should, as always, be multiplied by .6745 to obtain the probable error.

TABLE XLI—*Continued*

STANDARD AND PROBABLE ERRORS OF CERTAIN MEASURES ACCORDING TO CONVENTIONAL LARGE SAMPLE THEORY

| Measure | Standard Error ($\epsilon$) | Probable Error ($PE$) |
|---|---|---|
| Coefficient of rank correlation ($\rho$) | $(1 + .086\rho^2 + .013\rho^4 + .002\rho^6)\frac{1-\rho^2}{\sqrt{N}}$ | $.6745(1 + .086\rho^2 + .013\rho^4 + .002\rho^6)\frac{1-\rho^2}{\sqrt{N}}$ |
| Coefficient of correlation estimated from coefficient of rank correlation $\rho$ | $1.0472(1 + .042r^2 + .008r^4 + .002r^6)\frac{1-r^2}{\sqrt{N}}$ | $.7063(1 + .042r^2 + .008r^4 + .002r^6)\frac{1-r^2}{\sqrt{N}}$ |
| Coefficient of partial correlation ($r_{12\cdot3\ldots n}$) | $\frac{1-r^2_{12\cdot3\ldots n}}{\sqrt{N}}$ | $.6745\frac{1-r^2_{12\cdot3\ldots n}}{\sqrt{N}}$ |
| Coefficient of multiple correlation ($R_{1\cdot23\ldots n}$) | $\frac{1-R^2_{1\cdot23\ldots n}}{\sqrt{N}}$ | $.6745\frac{1-R^2_{1\cdot23\ldots n}}{\sqrt{N}}$ |
| Partial or multiple coefficient of regression ($b_{12\cdot3\ldots n}$) | $\frac{\sigma_{1\cdot23\ldots n}}{\sigma_{2\cdot34\ldots n}\sqrt{N}}$ | $.6745\frac{\sigma_{1\cdot23\ldots n}}{\sigma_{2\cdot34\ldots n}\sqrt{N}}$ |
| Ratio of correlation ($\eta$) | $\frac{1-\eta^2}{\sqrt{N}}$ | $.6745\frac{1-\eta^2}{\sqrt{N}}$ |
| Bi-serial coefficient of correlation ($r_{bis}$) | $\frac{\frac{\sqrt{pq}}{.3989h} - r^2_{bis}}{\sqrt{N}}$ | $.6745\frac{\frac{\sqrt{pq}}{.3989h} - r^2_{bis}}{\sqrt{N}}$ |

TABLE XLI—*Concluded*

STANDARD AND PROBABLE ERRORS OF CERTAIN MEASURES ACCORDING TO CONVENTIONAL LARGE SAMPLE THEORY

| Measure | Standard Error ($\epsilon$) | Probable Error ($PE$) |
|---|---|---|
| Coefficient of contingency ($C$) | $\frac{1}{\sqrt{N}}\sqrt{\frac{\frac{\frac{1}{N}\Sigma\left[\frac{\left(f_{rc}-\frac{f_r f_c}{N}\right)^3}{\left(\frac{f_r f_c}{N}\right)}\right]}{T-1}-T+2}{T^3}}$ | $\frac{.6745}{\sqrt{N}}\sqrt{\frac{\frac{\frac{1}{N}\Sigma\left[\frac{\left(f_{rc}-\frac{f_r f_c}{N}\right)^3}{\left(\frac{f_r f_c}{N}\right)^2}\right]}{T-1}-T+2}{T^3}}$ |
| * Coefficient of tetrachoric correlation ($r_{tet}$) | $\frac{\sqrt{\frac{(a+b)(a+c)(b+d)(c+d)}{N^4}}}{.1591hh'\sqrt{N}}\sqrt{\left[1-\left(\frac{\sin^{-1}r_{tet}}{90^\circ}\right)^2\right](1-r^2_{tet})}$ | |
| Frequency in a single class ($f$) | $\sqrt{\frac{f(N-f)}{N}}$ or $\sqrt{f-\frac{f^2}{N}}$ | $.6745\sqrt{\frac{f(N-f)}{N}}$ or $.6745\sqrt{f-\frac{f^2}{N}}$ |

* In these cases the standard error alone is given. It should, as always, be multiplied by .6745 to obtain the probable error.

or estimated parameters can be used instead of obtained statistics.

The one correction that can be made most frequently applies whenever $\sigma$ or $MdD$ appears. If $\Sigma d^2$ or $\Sigma fd^2$, as the case may be, is divided by $N - 1$ instead of by $N$, a better estimate of the standard deviation of the whole population is secured. If the number of cases is very small, still further correction is needed. As a compromise between exact accuracy and ease of computation, it has been suggested that $N$ be used when the number of cases exceeds 30, $N - 1$ when it is between 20 and 30, $N - 2$ when it is between 10 and 20, and $N - 3$ when it is less than 10. The writer, however, recommends using $N - 1$ for samples containing more than 30 cases as well as for those with fewer than that number.

Another serious limitat'on is that the formulas, as given, include more than sampling errors. If the data are not perfectly reliable—in other words, if a repetition of the procedure by which they were obtained would not yield corresponding data for the same individual cases—*errors of measurement* are also present. The formulas in Table XLI include both types of errors. If the coefficient of reliability of a set of data is known, an error of sampling may be secured by multiplying a formula in the table by the square root of that coefficient, or $\sqrt{r_{\text{II}}}$. The corresponding error of measurement may be found by multiplying the tabled formula by $\sqrt{1 - r_{\text{II}}}$. For example, the formula given for the standard error of the mean, $\epsilon_M$, is $\frac{\sigma}{\sqrt{N}}$; therefore the standard error of sampling of the mean is $\frac{\sigma\sqrt{r_{\text{II}}}}{\sqrt{N}}$ and the standard eror of measurement of the mean is

$$\frac{\sigma\sqrt{1 - r_{\text{II}}}}{\sqrt{N}}.$$

If the data are perfectly accurate, so that $r_{\text{II}} = +1.00$, then the error of measurement becomes 0 and that of sampling is given by the formula in Table XLI.

Another correction that makes for greater exactness is to substitute $N - n$ ($n$ being the number of variables or attri-

butes concerned) for $N$. Thus for a measure of central tendency or variability involving a single distribution, $N - 1$ is used; for one of simple correlation or regression involving two, $N - 2$; and so on. Making this correction amounts to dividing by the number of degrees of freedom instead of by the number of cases.

Most users of sampling error formulas overlook or neglect the fact that, as generally given and employed, they assume an infinite universe rather than take account of the number of cases in the universe. The number of cases in a universe is not always known, but when it is the formulas for errors of the mean and all others that contain $\frac{\sigma}{\sqrt{N}}$ or any multiple thereof should be multiplied by

$$\sqrt{1 - \frac{n-1}{N-1}},$$

in which $n$ is the number of cases in the sample and $N$ that in the universe. A close approximation may be obtained by employing

$$\sqrt{1 - \frac{n}{N}},$$

which is sometimes written $\sqrt{1 - p}$, $p$ being equal to $\frac{n}{N}$ or, in other words, being the per cent that the sample is of the universe. Inspection of the formula shows that the larger the per cent of the universe included in a sample, the smaller the error of sampling. For example, if a sample contains 1 per cent of the cases in a universe, the multiplier is approximately .995; if it contains 2 per cent, it is .99; if 5 per cent, .975; if 10 per cent, .95; if 25 per cent, .87; if 50 per cent, .71; if 75 per cent, .50; and so on until, if it contains 100 per cent, the multiplier is .00, since then there can be no error of sampling.

The formulas for errors in averages overestimate the sizes of such errors if distributions of data are leptokurtic and underestimate them if they are platykurtic. The same is true of all error formulas given in Table XLI with regard to the shapes of the distributions of the errors themselves rather than of the data to which they apply.

## Computing and interpreting errors of sampling

To illustrate the computation and interpretation of errors of sampling, some of the data from the often-used set of 44 cases may be considered as a sample of a larger number and may be employed. For them, $M = 253.64$ and $\sigma = 52.96$. Therefore, according to its conventional formula, $\frac{\sigma}{\sqrt{N}}$,

$$\epsilon_M = \frac{52.96}{\sqrt{44}} = 7.98.$$

If, as suggested above, $N - 1$ is used instead of $N$,[5]

$$\epsilon_M = \frac{52.96}{\sqrt{43}} = 8.08.$$

If the coefficient of reliability of the data were known, the value just given should be multiplied by $\sqrt{r_{11}}$. Since it is not known, it will be assumed to be $+.90$. Employing this value, we have $8.08\sqrt{.90} = 7.67$. If we further assume that the 44 cases constitute 5 per cent of the total group, we should multiply 7.67 by $\sqrt{1. - .05}$, which gives a result of 7.47, a better value than any of those preceding it.

As previously suggested, this or any other such error may be interpreted, as may any similar deviation, by the use of the appropriate one of the tables in the Appendix. That is, the standard error of any statistic of a random sample is merely the standard deviation of a distribution of that statistic for an infinite number of random samples of the same size. Since in this case $\epsilon_M = 7.47$, there are approximately 68.27 to 31.73, or 2.15 to 1, chances that the mean actually obtained is within 7.47 of the true mean of the whole population from which the sample was drawn; 95.45 to 4.55, or 21 to 1, chances that the obtained mean is within $2 \times 7.47$, or 14.94, of the true mean; and so on for any multiple or fraction of $\epsilon_M$.

[5] $N - 1$ should be substituted for $N$ in the calculation of $\sigma$ rather than in that of $\epsilon$. Since, however, the result is the same whether $\sigma$ is computed with the use of $N$ and then $\epsilon$ with $N - 1$, or $\sigma$ with $N - 1$ and $\epsilon$ with $N$, it is customary to introduce $N - 1$ with $\epsilon$. If it is used in computing $\sigma$ of the data employed, $\sigma = 53.58$ rather than 52.96 and

$$\epsilon_M = \frac{53.58}{\sqrt{44}} = 8.08,$$

as above.

Sometimes a somewhat-different approach is convenient. What is desired may be what per cent of obtained measures—in this instance, means—is within a specified number of points of the true mean. This may also be stated as the chances that any one of the obtained measures is within the same number of points of the true mean. To do this, the number of points is divided by the standard error and the result, which is merely the distance expressed in standard error or deviation units, is looked up in the table. If the distance is 10 points, for example, 10 is divided by 7.47, the result being 1.34. According to the table, 2 × .4096, or about 82 per cent, of the area of a normal curve is within $1.34\epsilon$ of the mean; hence 82 per cent of the obtained means will probably be within 10 points of the true mean. Stating it otherwise, the chances are 82 to 18, or about 4.6 to 1, that any one obtained mean is within that distance.

Still different is the situation in which we wish to know the chances that the true mean is as small or as large as a stated value. We may, for example, want to know the chances that, for the data just employed, the true mean is not 270 or larger. The procedure is to divide the difference between the true and the obtained means by its standard or probable error and find the last-column entry corresponding to the result. In this case, $270 - 253.64 = 16.36$, and this divided by $\epsilon_M$, 7.47, gives 2.19. The last-column entry corresponding to this in the first column of the standard deviation table is approximately 69; hence there are 69 chances to 1 that the true mean would not be 270 or larger.

Similarly, by the use of either the standard or the probable error, the reliability of other measures when they are regarded as statistics or sampling measures may be determined and expressed. The practice of stating the standard or probable error with the measure to which it refers should be much more frequent than it is. When the probable error is so stated, it is conventionally written after the measure with a plus-or-minus sign between them. For example, $PE_M$ for the data just used is $.6745 \times 7.47 = 5.04$; hence the mean may be written $253.64 \pm 5.04$. There is no corresponding generally accepted method for the standard error.

## The significance of differences

The reliability or significance of differences is frequently important, particularly in educational experimentation. Such questions as whether a difference in the achievements of two groups wherein different methods were employed indicates a real difference or probably is due to chance, whether a single pupil's gain in weight or speed of reading or ability to solve problems in arithmetic is significant or not, and so on, often arise. Sometimes not merely differences but differences between differences are involved.

The general formula for the standard error of a sum or difference is

$$\epsilon_{1\pm 2} = \sqrt{\epsilon_1^2 + \epsilon_2^2 \pm 2r_{12}\epsilon_1\epsilon_2}\,,$$

in which the subscripts "1" and "2" represent the two measures the sum or difference of which is in question. If a sum is involved, the last term is preceded by a plus sign; if a difference, by a minus sign. If there is no correlation between the measures concerned, the last term becomes zero, leaving merely

$$\epsilon_{1\pm 2} = \sqrt{\epsilon_1^2 + \epsilon_2^2}.$$

All too often, however, workers have employed this short form when correlation is present and the long form should be used. The effect of so doing is to yield too small values when sums are concerned and too large ones when differences are, with consequent interpretations of too high reliability of sums and too low reliability of differences.

Since, in actual practice, the reliability of sums is desired far less often than that of differences, only the latter will be illustrated. Let us suppose that we wish to determine the significance of the gain in mean ability to solve a specified type of arithmetic examples resulting from practice given a group of 34 pupils, and that their mean score before the practice was 14.2; that after it, 16.4; the corresponding standard deviations, 3.1 and 3.8; and the coefficient of correlation between the two series of scores, +.73. First, the standard errors of the means

must be found. Using the subscript "1" to refer to the scores before the practice and "2" to refer to those after, we have

$$\epsilon_{M_1} = \frac{3.1}{\sqrt{33}} = .54 \text{ and } \epsilon_{M_2} = \frac{3.8}{\sqrt{33}} = .66.$$

Therefore

$$\epsilon_{M_1-M_2} = \sqrt{.54^2 + .66^2 - 2. \times .73 \times .54 \times .66^{6}} = .45.$$

The difference between means is $16.4 - 14.2 = 2.2$, which divided by .45 gives 4.89, the size of the difference in terms of its own standard error. This quotient is sometimes called the *significance ratio.* The last column of the standard deviation table in the Appendix gives the chances to 1 that the difference is real—in other words, that the true mean, or mean of the universe, after the practice is greater than the mean before the practice. For $\sigma$ or $\epsilon$ equal to 4.89, the chances are approximately 2,000,000 to 1.

The formula for the error of the sum or difference of two sums or differences is the same as that for a simple sum or difference, except that the terms under the radical are those for sums or differences rather than for simple measures. Thus, employing the subscripts "1" and "2" for two measures of which differences are taken and "3" and "4" for two others,

$$\epsilon_{(1-2)-(3-4)} = \sqrt{\epsilon_{1-2}^2 + \epsilon_{3-4}^2 - 2r_{(1-2)(3-4)}\,\epsilon_{1-2}\epsilon_{3-4}}\,.$$

This may be expanded and written in terms of standard errors of simple measures rather than of differences by substituting for $\epsilon_{1-2}$ and $\epsilon_{3-4}$ their values, which are, of course,

$$\epsilon_{1-2} = \sqrt{\epsilon_1^2 + \epsilon_2^2 - 2r_{12}\epsilon_1\epsilon_2} \text{ and } \epsilon_{3-4} = \sqrt{\epsilon_3^2 + \epsilon_4^2 - 2r_{34}\epsilon_1\epsilon_2}.$$

To illustrate such a case, let us assume that for another group, containing 39 pupils, given a different type of practice, the means before and after were 13.6 and 15.3; the standard deviations, 2.9 and 3.7; and the coefficient of correlation, +.68.

[6] Although the coefficient of correlation given in the text—.73—is that between the series of scores, and the formula for $\epsilon_{M_1-M_2}$ calls for that between the means, the former is used because it can be shown that the two are the same.

Using the subscript "3" to refer to measures before the practice and "4" to those after, we have

$$\epsilon_{M_3} = \frac{2.9}{\sqrt{38}} = .47 \text{ and } \epsilon_{M_4} = \frac{3.7}{\sqrt{38}} = .60;$$

whence

$$\epsilon_{M_3-M_4} = \sqrt{.47^2 + .60^2 - 2. \times .68 \times .47 \times .60} = .44.$$

Since there cannot be correlation between two different groups,

$$\epsilon_{(M_1-M_2)-(M_3-M_4)} = \sqrt{\epsilon^2_{(M_1-M_2)} + \epsilon^2_{(M_3-M_4)}} = \sqrt{.45^2 + .44^2} = .63.$$

According to the suggestion at the end of the last paragraph, the same result can also be obtained thus:

$$\sqrt{\epsilon^2_{M_1} + \epsilon^2_{M_2} + \epsilon^2_{M_3} + \epsilon^2_{M_4} - 2r_{M_1M_2}\epsilon_{M_1}\epsilon_{M_2} - 2r_{M_3M_4}\epsilon_{M_3}\epsilon_{M_4}} =$$
$$\sqrt{.54^2+.66^2+.47^2+.60^2-2.\times.73\times.54\times.66-2.\times.68\times.47\times.60} = .63.$$

The difference between differences is $(16.4 - 14.2) - (15.3 - 13.6) = 2.2 - 1.7 = .5$. Dividing this by its standard error, .63, we get .79, the difference between differences expressed in terms of its standard error. For this value in the first column of Table XLIV the corresponding entry in the last column is, by interpolation, approximately 3.7; hence there are that many chances to 1 that there is a real difference in the direction indicated, that is, that the gain by the first type of practice is larger than that by the second.

The name *critical ratio* has been given to the result secured by dividing a difference between means by its probable error. Sometimes also

$$CR = \frac{\text{difference}}{\epsilon_{diff}},$$

and sometimes it is employed to refer to differences between other measures than means. The writer recommends that it be used only as

$$\frac{\text{difference}}{PE_{diff}},$$

but in connection with all differences.

## The experimental coefficient method

A method of interpreting the significance of differences that has received considerable use is that based upon the *experimental coefficient.* This is defined by the formula

$$EC = \frac{\text{difference}}{2.78\epsilon_{diff}} \text{ or } \frac{\text{difference}}{4.12PE_{diff}}.$$

The multiplier in the denominator was so chosen that when $EC = 1.00$ the chances that a difference is significant are 369 to 1, which McCall, who suggested the $EC$, called practical certainty. This is at variance with the more usual definition of practical certainty as 99 chances to 1, or sometimes as 999 to 1. The present writer strongly objects to the use of any so defined critical point and the consequent dichotomous classification of differences as definitely either significant or not significant, certain or not certain, reliable or not reliable. It seems absurd to consider 99 chances to 1 as certainty, but not 98 to 1. Instead, the actual chances should be stated in each case. Table XLII gives chances of significance for various other values of the experimental coefficient. The

TABLE XLII

CHANCES THAT A DIFFERENCE IS SIGNIFICANT CORRESPONDING TO VARIOUS VALUES OF THE EXPERIMENTAL COEFFICIENT

| Experimental Coefficient | Approximate Chances to 1 |
|---|---|
| .1 | 1.6 |
| .2 | 2.5 |
| .3 | 4.0 |
| .4 | 6.5 |
| .5 | 11. |
| .6 | 20. |
| .7 | 38. |
| .8 | 76. |
| .9 | 160. |
| 1.0 | 370 |
| 1.1 | 900 |
| 1.2 | 2400 |
| 1.3 | 6700 |
| 1.4 | 20000 |
| 1.5 | 67000 |

expression *chances of significance* in this connection means just the same as in the last section—chances that the difference really is in the indicated direction.

The experimental coefficient method gives just the same chances as that presented in the last section, but requires the additional step involved by introducing the factor 2.78 or 4.12, hence seems to the writer less deserving of use than the other method. In the last example of that section, for example,

$$EC = \frac{.5}{2.78 \times .63} = .28.$$

For this value, Table XLII yields chances of about 3.7 to 1, the same as were already determined.

## Small sample theory

The most commonly applied small sample theory is that which makes use of $t$ instead of $\epsilon$ or $PE$. Originally the use of $t$ was limited to situations involving differences between means and its formula was

$$t = \frac{M - \tilde{M}}{\tilde{\epsilon}_M}.$$

Later, however, it came to be employed in connection with other differences also. It may be defined, therefore, as the difference between a statistic and the corresponding parameter divided by the estimated standard error of the measure concerned. Its distribution is leptokurtic in shape, although as the number of cases becomes large it approaches normality; hence it requires a special table for its interpretation in terms of probability or significance. For large numbers of cases it yields interpretations quite similar to those obtained by large sample theory; for very small numbers it yields markedly different interpretations.

Values corresponding to those in the last four columns of a $\sigma$ or $\epsilon$ table of the normal curve may be found for $t$. Since, however, they differ according to the number of degrees of freedom, they require much more elaborate tables. In default of these, condensed tables for $t$ are given in several forms. Most of them give values for only certain stated levels of

significance or critical points. The most frequent are chances of 20 to 1, often called *significant*, and of 100 to 1, called *very significant*. This practice is one that the writer deplores. Instead, he prefers a table containing entries corresponding to values of $t$ and numbers of degrees of freedom. The entries may be similar to those in any one of the four columns mentioned above. Probably entries which correspond to those in the next-to-the-last column—that is, show the ratio of area within a distance of $t$ from the mean to the area farther than that—are most convenient. In other words, they show the chances to 1 that a statistic does not differ more than $t$ from the corresponding parameter. Just as for a normal distribution, so here the chances that a statistic is as far or farther away from the parameter and in the same direction as the difference equals 1 more than twice the chances mentioned in the preceding sentence.

Table XLIII has been prepared for use in this connection. It gives, for certain values of $t$ and numbers of degrees of freedom, the chances to 1 that a statistic does not differ from the corresponding parameter by more than the value of $t$. The entries are accurate to two significant figures only and, to economize space, are not given when 100 or greater. Its use will be illustrated in later paragraphs.

In actual use $t$ is often employed to find the chance that, if an assumed or hypothetical value of a measure is the true one, an actually obtained one would be found. For example, if the mean of the population of which the previously used distribution of 44 cases is a sample is 260, what is the chance that an obtained mean would be as far away as is 253.64? $260 - 253.64 = 6.36$, which divided by 8.08, the value already obtained for estimated $\tilde{\epsilon}_M$, gives $t = .79$. For $t = .79$ and 11 degrees of freedom the interpolated tabular entry is 1.27. In other words, the chances are 1.27 to 1 that an obtained mean will not be less than 253.64 if the true mean is 260.

For comparison, results by large sample theory will be found. The formula for the standard error of a difference applies. Thus

$$\epsilon_{M-\tilde{M}} = \sqrt{\epsilon_M^2 + \tilde{\epsilon}_M^2},$$

# TABLE XLIII

CHANCES TO 1 THAT A STATISTIC DOES NOT DIFFER FROM THE CORRESPONDING PARAMETER BY MORE THAN $t$

| $t$ | Number of Degrees of Freedom | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | ∞ |
| .0 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 |
| .1 | .07 | .08 | .08 | .08 | .08 | .08 | .08 | .08 | .08 | .08 | .08 | .08 | .08 | .08 | .09 | .09 | .09 | .09 | .09 | .09 | .09 | .09 | .09 | .09 | .09 | .09 |
| .2 | .14 | .16 | .17 | .17 | .18 | .18 | .18 | .18 | .18 | .18 | .18 | .18 | .18 | .18 | .18 | .18 | .19 | .19 | .19 | .19 | .19 | .19 | .19 | .19 | .19 | .19 |
| .3 | .23 | .26 | .28 | .28 | .29 | .29 | .30 | .30 | .30 | .30 | .30 | .30 | .30 | .30 | .30 | .30 | .30 | .30 | .30 | .30 | .30 | .30 | .30 | .30 | .30 | .31 |
| .4 | .32 | .37 | .40 | .41 | .42 | .43 | .43 | .43 | .43 | .44 | .44 | .44 | .44 | .44 | .44 | .44 | .44 | .44 | .44 | .44 | .44 | .44 | .44 | .44 | .44 | .45 |
| .5 | .42 | .50 | .54 | .56 | .57 | .58 | .58 | .59 | .59 | .59 | .59 | .60 | .60 | .60 | .60 | .60 | .60 | .60 | .61 | .61 | .61 | .61 | .61 | .61 | .61 | .62 |
| .6 | .52 | .64 | .70 | .72 | .74 | .75 | .76 | .77 | .77 | .78 | .78 | .79 | .79 | .79 | .79 | .80 | .80 | .80 | .80 | .80 | .80 | .80 | .80 | .81 | .81 | .82 |
| .7 | .64 | .80 | .87 | .91 | .94 | .96 | .97 | .98 | .99 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.1 |
| .8 | .75 | .97 | 1.1 | 1.1 | 1.2 | 1.2 | 1.2 | 1.2 | 1.3 | 1.3 | 1.3 | 1.3 | 1.3 | 1.3 | 1.3 | 1.3 | 1.3 | 1.3 | 1.3 | 1.3 | 1.3 | 1.3 | 1.3 | 1.3 | 1.3 | 1.4 |
| .9 | .87 | 1.2 | 1.3 | 1.4 | 1.4 | 1.5 | 1.5 | 1.5 | 1.6 | 1.6 | 1.6 | 1.6 | 1.6 | 1.6 | 1.6 | 1.6 | 1.6 | 1.6 | 1.6 | 1.6 | 1.6 | 1.6 | 1.6 | 1.7 | 1.7 | 1.7 |
| 1.0 | 1.0 | 1.4 | 1.6 | 1.7 | 1.8 | 1.8 | 1.9 | 1.9 | 1.9 | 1.9 | 2.0 | 2.0 | 2.0 | 2.0 | 2.0 | 2.0 | 2.0 | 2.0 | 2.0 | 2.0 | 2.0 | 2.0 | 2.1 | 2.1 | 2.1 | 2.2 |
| 1.1 | 1.1 | 1.6 | 1.8 | 2.0 | 2.1 | 2.2 | 2.2 | 2.3 | 2.3 | 2.4 | 2.4 | 2.4 | 2.4 | 2.4 | 2.5 | 2.5 | 2.5 | 2.5 | 2.5 | 2.5 | 2.5 | 2.5 | 2.5 | 2.5 | 2.5 | 2.7 |
| 1.2 | 1.3 | 1.8 | 2.2 | 2.4 | 2.5 | 2.6 | 2.7 | 2.8 | 2.8 | 2.9 | 2.9 | 3.0 | 3.0 | 3.0 | 3.0 | 3.0 | 3.1 | 3.1 | 3.1 | 3.1 | 3.1 | 3.1 | 3.1 | 3.1 | 3.1 | 3.3 |
| 1.3 | 1.4 | 2.1 | 2.5 | 2.8 | 3.0 | 3.1 | 3.3 | 3.4 | 3.4 | 3.5 | 3.6 | 3.6 | 3.6 | 3.7 | 3.7 | 3.7 | 3.7 | 3.8 | 3.8 | 3.8 | 3.8 | 3.8 | 3.9 | 3.9 | 3.9 | 4.2 |
| 1.4 | 1.5 | 2.4 | 2.9 | 3.3 | 3.5 | 3.7 | 3.9 | 4.0 | 4.1 | 4.2 | 4.3 | 4.4 | 4.4 | 4.5 | 4.5 | 4.5 | 4.6 | 4.6 | 4.6 | 4.7 | 4.7 | 4.7 | 4.7 | 4.8 | 4.8 | 5.2 |
| 1.5 | 1.7 | 2.7 | 3.3 | 3.8 | 4.2 | 4.4 | 4.6 | 4.8 | 5.0 | 5.1 | 5.2 | 5.3 | 5.3 | 5.4 | 5.5 | 5.5 | 5.6 | 5.6 | 5.7 | 5.7 | 5.7 | 5.8 | 5.8 | 5.8 | 5.9 | 6.5 |
| 1.6 | 1.8 | 3.0 | 3.8 | 4.4 | 4.9 | 5.2 | 5.5 | 5.7 | 5.9 | 6.1 | 6.2 | 6.4 | 6.5 | 6.6 | 6.6 | 6.7 | 6.8 | 6.9 | 6.9 | 7.0 | 7.0 | 7.1 | 7.1 | 7.2 | 7.2 | 8.1 |
| 1.7 | 2.0 | 3.3 | 4.3 | 5.1 | 5.7 | 6.1 | 6.5 | 6.8 | 7.1 | 7.3 | 7.5 | 7.7 | 7.8 | 8.0 | 8.1 | 8.2 | 8.3 | 8.4 | 8.5 | 8.6 | 8.6 | 8.7 | 8.8 | 8.9 | 8.9 | 10. |
| 1.8 | 2.1 | 3.7 | 4.9 | 5.8 | 6.6 | 7.2 | 7.7 | 8.1 | 8.5 | 8.8 | 9.1 | 9.3 | 9.5 | 9.7 | 9.9 | 10. | 10. | 10. | 10. | 10. | 11. | 11. | 11. | 11. | 11. | 13. |
| 1.9 | 2.2 | 4.1 | 5.5 | 6.7 | 7.6 | 8.4 | 9.1 | 9.6 | 10. | 11. | 11. | 11. | 12. | 12. | 12. | 12. | 12. | 13. | 13. | 13. | 13. | 13. | 13. | 14. | 14. | 16. |
| 2.0 | 2.4 | 4.4 | 6.2 | 7.6 | 8.8 | 9.8 | 11. | 11. | 12. | 13. | 13. | 14. | 14. | 14. | 15. | 15. | 15. | 15. | 16. | 16. | 16. | 16. | 17. | 17. | 17. | 21. |
| 2.1 | 2.5 | 4.9 | 6.9 | 8.7 | 10. | 11. | 13. | 13. | 14. | 15. | 16. | 16. | 17. | 17. | 18. | 18. | 19. | 19. | 19. | 20. | 20. | 20. | 20. | 21. | 21. | 27. |
| 2.2 | 2.7 | 5.3 | 7.7 | 9.8 | 12. | 13. | 15. | 16. | 17. | 18. | 19. | 20. | 21. | 21. | 22. | 22. | 23. | 23. | 24. | 24. | 25. | 25. | 25. | 26. | 26. | 35. |
| 2.3 | 2.8 | 5.8 | 8.5 | 11. | 13. | 15. | 17. | 19. | 20. | 22. | 23. | 24. | 25. | 26. | 27. | 27. | 28. | 29. | 29. | 30. | 30. | 31. | 32. | 32. | 33. | 46. |
| 2.4 | 3.0 | 6.2 | 9.4 | 12. | 15. | 18. | 20. | 22. | 24. | 26. | 27. | 29. | 30. | 31. | 33. | 33. | 34. | 35. | 36. | 37. | 38. | 39. | 40. | 41. | 42. | 60. |
| 2.5 | 3.1 | 6.7 | 10. | 14. | 17. | 21. | 23. | 26. | 28. | 31. | 33. | 35. | 37. | 38. | 40. | 41. | 42. | 44. | 45. | 46. | 48. | 49. | 50. | 52. | 53. | 80. |
| 2.6 | 3.3 | 7.2 | 11. | 16. | 20. | 24. | 27. | 31. | 33. | 37. | 40. | 42. | 44. | 47. | 49. | 51. | 53. | 55. | 56. | 57. | 59. | 60. | 62. | 63. | 65. | * |
| 2.7 | 3.4 | 7.8 | 13. | 18. | 22. | 27. | 32. | 36. | 40. | 44. | 48. | 51. | 54. | 57. | 60. | 62. | 65. | 67. | 69. | 71. | 74. | 76. | 78. | 80. | 82. | * |

* The blank spaces represent entries of 100 or more.

TABLE XLIII—*Concluded*

CHANCES TO 1 THAT A STATISTIC DOES NOT DIFFER FROM THE CORRESPONDING PARAMETER BY MORE THAN $t$

| $t$ | Number of Degrees of Freedom | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 |
| 2.8 | 3.6 | 8.3 | 14. | 19. | 25. | 31. | 37. | 42. | 47. | 52. | 57. | 62. | 66. | 70. | 74. | 77. | 80. | 84. | 87. | 90. | 93. | 97. | * | | |
| 2.9 | 3.7 | 8.9 | 15. | 22. | 29. | 35. | 42. | 50. | 56. | 62. | 68. | 74. | 80. | 85. | 90. | 95. | 99. | * | | | | | | | |
| 3.0 | 3.9 | 9.5 | 16. | 24. | 32. | 41. | 49. | 58. | 66. | 74. | 82. | 90. | 97. | * | | | | | | | | | | | |
| 3.1 | 4.0 | 10. | 18. | 27. | 36. | 46. | 56. | 67. | 77. | 88. | * | | | | | | | | | | | | | | |
| 3.2 | 4.2 | 11. | 19. | 29. | 41. | 53. | 66. | 78. | 92. | * | | | | | | | | | | | | | | | |
| 3.3 | 4.3 | 11. | 21. | 32. | 46. | 60. | 75. | 92. | * | | | | | | | | | | | | | | | | |
| 3.4 | 4.5 | 12. | 23. | 36. | 51. | 68. | 87. | * | | | | | | | | | | | | | | | | | |
| 3.5 | 4.7 | 13. | 24. | 39. | 57. | 77. | 99. | * | | | | | | | | | | | | | | | | | |
| 3.6 | 4.8 | 13. | 26. | 43. | 63. | 87. | * | | | | | | | | | | | | | | | | | | |
| 3.7 | 5.0 | 14. | 28. | 47. | 70. | 99. | * | | | | | | | | | | | | | | | | | | |
| 3.8 | 5.1 | 15. | 30. | 51. | 78. | * | | | | | | | | | | | | | | | | | | | |
| 3.9 | 5.3 | 16. | 32. | 56. | 87. | * | | | | | | | | | | | | | | | | | | | |
| 4.0 | 5.4 | 16. | 35. | 61. | 97. | | | | | | | | | | | | | | | | | | | | |
| 4.1 | 5.6 | 17. | 37. | 67. | * | | | | | | | | | | | | | | | | | | | | |
| 4.2 | 5.7 | 18. | 40. | 73. | * | | | | | | | | | | | | | | | | | | | | |
| 4.3 | 5.9 | 19. | 42. | 79. | * | | | | | | | | | | | | | | | | | | | | |
| 4.4 | 6.0 | 20. | 45. | 85. | * | | | | | | | | | | | | | | | | | | | | |
| 4.5 | 6.2 | 21. | 48. | 92. | * | | | | | | | | | | | | | | | | | | | | |
| 4.6 | 6.3 | 22. | 51. | 99. | * | | | | | | | | | | | | | | | | | | | | |
| 4.7 | 6.5 | 23. | 54. | * | | | | | | | | | | | | | | | | | | | | | |
| 4.8 | 6.6 | 24. | 57. | * | | | | | | | | | | | | | | | | | | | | | |
| 4.9 | 6.8 | 25. | 61. | * | | | | | | | | | | | | | | | | | | | | | |
| 5.0 | 7.0 | 25. | 64. | * | | | | | | | | | | | | | | | | | | | | | |
| 6.0 | 8.5 | 37. | * | | | | | | | | | | | | | | | | | | | | | | |
| 7.0 | 10. | 51. | * | | | | | | | | | | | | | | | | | | | | | | |
| 8.0 | 12. | 66. | * | | | | | | | | | | | | | | | | | | | | | | |
| 9.0 | 13. | 85. | * | | | | | | | | | | | | | | | | | | | | | | |
| 10.0 | 15. | * | | | | | | | | | | | | | | | | | | | | | | | |

* The blank spaces represent entries of 100 or more.

but, since $\tilde{M}$ has no error, this becomes

$$\sqrt{\epsilon_M^2 + 0^2} = \sqrt{\epsilon_M^2} = \epsilon_M.$$

Therefore the difference divided by its standard error equals

$$\frac{M - \tilde{M}}{\epsilon_M},$$

which is the same as $t$ or .79. The interpolated entry in the next-to-the-last column of the first table in the Appendix for $\epsilon = .79$ is 1.33 and that in the last column is 3.66. These do not differ greatly from the corresponding values of 1.27 and 3.54 found by the use of $t$. In general, the smaller the values of $t$ and of the difference divided by $\epsilon$, the less do the chances based upon them differ, and vice versa.

The procedure based upon $t$ may also be employed to determine chances of difference between two obtained means or other measures as well as between one obtained and one assumed true one. In this case

$$t = \frac{M_1 - M_2}{\epsilon_{M_1 - M_2}},$$

that is, the difference between the means divided by the standard error of that difference. Applying it to the data used on page 242, we have

$$t = \frac{16.4 - 14.2}{.45} = 4.89.$$

For this value of $t$ and 33 degrees of freedom the entry that would be in Table XLII if it were extended is evidently far larger than 100, which indicates practical certainty as to the difference being significant.

## Degrees or levels of confidence

The use of degrees or levels of confidence in stating the significance of results has been employed increasingly in recent years. Although the present writer does not approve it, he will explain it briefly. A *degree* or *level of confidence* is the per cent of chances that a result is due to chance rather than to dependable causes: For example, if there are 90 chances to 10 in favor of dependability, the level of confidence is 10 per cent;

if 97 to 3, it is 3 per cent; and so on. Tables for the interpretation of sampling errors may give values at given levels of confidence, such as the 5 per-cent, the 2 per-cent, and the 1 per-cent level. Such tables may then be used, more or less in advance of the determination of measures, to indicate their sizes for certain degrees of confidence.

To illustrate the usage just mentioned, we may employ the first data used on page 242 in connection with the difference of two means. The standard error of that difference, $\epsilon_{M_1-M_2}$, was found to be .45. Therefore, if the difference is at the 5 per-cent level of confidence, it must be as many times its standard error as corresponds to chances of 95 to 5, or a ratio of 19 to 1. By interpolation, a $\sigma$ distance of 1.64 is found to correspond to an entry of 19 in the last column of the standard deviation table in the Appendix. Since $1.64 \times .45 = .74$, the difference between the two means concerned must be .74 in order that there be 95 chances to 5, or 19 to 1, that the one having the larger obtained value also has the larger true value. In other words, the 5 per-cent level of confidence in this situation requires a difference of .74. Similarly the 1 per-cent level, or 99 chances to 1, requires a difference of $2.32 \times .45 = 1.04$. Another way of stating these facts is that if the difference is .74 there are 5 chances out of 100 that the larger obtained mean is larger because of errors of sampling and that if it is 1.04 there is 1 chance out of 100 that this is true.

## The null hypothesis

Sometimes the approach is through what is known as the *null hypothesis*. This was originally employed to designate any exact hypothesis which one may wish to disprove, but is frequently limited in use to situations in which the hypothesis is that the parameter in question is zero. If this hypothesis is rejected, the statistic is significant; that is, it signifies that the parameter is not zero. In this connection the expression *level of significance* is employed in the same way as level of confidence in the other approach.

Situations in which the null hypothesis is appropriate are those in which the point at issue is the possibility that param-

eters are zero rather than the limits within which they lie. In such an instance as that used in the preceding section and previously, and others wherein differences are concerned, this condition is likely to hold. Interest centers in whether or not the larger obtained measure signifies the larger true measure rather than in the probable limits of the latter. In other words, the direction rather than the size of the difference is important. This is what was determined in the preceding section.

## Errors of statistics from matched samples

Not infrequently groups of pupils employed in experimental studies are matched. When this is done, the effect is the same as if an element of correlation were introduced; therefore, formulas for errors based on data secured from such groups must take account of this fact. Instead of using the ordinary error formulas unchanged, the multiplier $\sqrt{1 - r^2}$ should be introduced if the groups are matched on a fallible or unreliable criterion; $\sqrt{1 - r}$, if they are matched on an infallible or true criterion. In either case $r$ is the coefficient of correlation between the matching element for the successive groups and is found by assuming that each pair or more of matched cases, one from each of two or more groups, may be considered as scores made by the same individual. Means are more frequently concerned than other statistics, but the application of the procedure is not limited to them.

## The reliability of coefficients of correlation

Although the reliability of the coefficient of correlation is generally estimated by comparing it with its standard or probable error as found by the formula in Table XLI, this practice is less valid for $r$ than for most of the other measures for which error formulas are given in the same table. For low values of $r$ from large samples the results are approximately correct, but when $r$ is large or the sample small they are liable

to be seriously inaccurate. A better formula for $\epsilon_r$ than

$$\frac{1 - r^2}{\sqrt{N}} \text{ or } \frac{1 - r^2}{\sqrt{N - 1}}$$

is

$$\frac{1 - r^2_{corr}}{\sqrt{N - 1}}\left(1 + \frac{11r^2_{corr}}{4(N - 1)}\right),$$

in which $r_{corr}$ is the value given by the second formula on page 133. If $r = .50$ and $N = 20$,

$$r_{corr} = .49 \text{ and } \epsilon_r = \frac{1 - .49^2}{\sqrt{20 - 1}}\left(1 + \frac{11 \times .49^2}{4(20 - 1)}\right) = .18,$$

which is slightly larger than the .17 which the formula from Table XLI gives.

A different approach sometimes useful is by means of Table XLIII. If $t$ is found by the formula

$$t = \sqrt{\frac{r^2(N - 2)}{1 - r^2}},$$

the corresponding entry in Table XLIII is the probability that if the true correlation is zero a value as large as the absolute value of the obtained $r$, or larger, would be secured by pure chance. For example, if $r = .50$ for 20 cases,

$$t = \sqrt{\frac{.50^2(20 - 2)}{1 - .50^2}} = 2.45.$$

The interpolated entry in Table XLIII corresponding to $t = 2.45$ and 18 degrees of freedom is approximately 40. In other words, the chances are 40 to 1 that a coefficient exceeding $\pm.50$ would not be secured if the true value of $r$ were .00. The chances against a value of $+.50$ or more are $2 \times 40 + 1$, or 81, to 1.

Except when $N$ is quite small, the formula

$$\epsilon_r = \frac{1}{\sqrt{N - 1}},$$

which is that for the standard error of $r$ when true $r = .00$, may be used for the same purpose. In this case it gives

$$\epsilon_r = \frac{1}{\sqrt{20 - 1}} = .229.$$

Dividing $r$, .50, by this gives 2.18, the value of $r$ in terms of its standard error. Using this as a first-column entry in the first table in the Appendix, we find a corresponding last-column entry of about 69, the chances to 1 that a value of +.50 or more would not be obtained. The discrepancy between 69 and 81 is due largely to there being only 20 cases concerned.

More elaborate and exact procedures for estimating the reliability of a coefficient of correlation have been proposed, but do not appear appropriate for presentation in an elementary text. Probably the best involves changing $r$ into another statistic, $z$, which has an approximately normal distribution although $r$ does not.

## EXERCISES AND PROBLEMS

Since the errors of almost all the measures called for in the exercises and problems at the ends of previous chapters can be found, similar computations will not be asked for here. It is recommended that the standard and probable errors of some of each type of measure for which error formulas are available be found, both by the formulas in Table XLI and with the appropriate corrections and modifications thereto suggested in the text. The exercises below call for certain applications of these formulas and uses of the procedures presented later in the chapter.

**1.** If $M = 48.6$, $\sigma = 5.4$, and $N = 65$, what is the chance that $M$ is

(a) within 2 points of $\tilde{M}$?
(b) within 10 points of $\tilde{M}$?
(c) more than 5 points from $\tilde{M}$?
(d) at least 3 points larger than $\tilde{M}$?
(e) at least 8 points smaller than $\tilde{M}$?

**2.** If $Q_1 = 64.0$, $Q_3 = 82.4$, and $N = 230$, what is the chance that
(a) $Q_1$ and $Q_3$ differ from their respective parameters by less than 4 points?
(b) $Q_1$ and $Q_3$ differ from their respective parameters by more than 6 points?
(c) $Q_1$ is at least 3 points more than $\tilde{Q}_1$?
(d) $Q_3$ is at least 8 points less than $\tilde{Q}_3$?
(e) $Q$ is within 2 points of $\tilde{Q}$?

3. If an experimental group of 33 pupils has a mean score of 74 on an initial test and of 97 on a final test, with accompanying standard deviations of 10.2 and 13.6, respectively, and a coefficient of correlation of .72; and a control group of the same size has corresponding means of 72 and 92, with standard deviations of 9.6 and 12.8, and a coefficient of .76; what is the chance that the greater gain of the experimental group is significant?

4. If the following data were obtained from an experiment, what is the chance that the difference in gains is significant?

| | | Initial | | Final | | |
|---|---|---|---|---|---|---|
| | $N$ | $Md$ | $Q$ | $Md$ | $Q$ | $r$ |
| Experimental | 40 | 35 | 6.3 | 64 | 8.7 | .68 |
| Control | 133 | 38 | 7.2 | 62 | 8.8 | .73 |

5. Find the chances for the data in Exercises 3 and 4 by the experimental coefficient method.

6. Find the chances for the data in Exercise 1 if $N$ is 12.

7. Find the chances for the data in Exercise 2 if $N$ is 16.

8. Find the chances for the data in Exercises 3 and 4 by small sample procedure if both values of $N$ in Exercise 3 are 10 and if its values in Exercise 4 are 8 and 15, respectively.

9. Find the chances for the data in Exercise 3 if the data are matched

(a) on a fallible basis with $r = .84$.
(b) on an infallible basis with $r = .96$.

10. If $r = .71$ for a sample of 85 cases, what is the chance that it

(a) differs from $\tilde{r}$ by more than .10?
(b) differs from $\tilde{r}$ by more than .04?
(c) differs from $\tilde{r}$ by less than .02?
(d) is at least .03 smaller than $\tilde{r}$?
(e) is at least .06 larger than $\tilde{r}$?

11. If $r = .32$ and $N = 17$, what are the chances that $\tilde{r}$ is not .00?

12. If $r = -.27$ and $N = 40$, what are the chances that $\tilde{r}$ is not .00?

## REFERENCES

Croxton, Frederick E., and Cowden, Dudley J., *Applied General Statistics*. Pages 325–331, 681–682, 875. New York: Prentice-Hall, Inc., 1939.

Ezekiel, Mordecai, *Methods of Correlation Analysis.* Second Edition. Chaps. II, XVIII–XX; pages 469–474. New York: John Wiley & Sons, Inc., 1941.

Fisher, R. A., *Statistical Methods for Research Workers.* Eighth Edition. Pages 112–140, 185–197. Edinburgh: Oliver and Boyd, 1941.

Garrett, Henry E., *Statistics in Psychology and Education.* Second Edition. Chap. VIII; pages 280–283. New York: Longmans, Green & Company, Inc., 1937.

Guilford, J. P., *Fundamental Statistics in Psychology and Education.* Chap. IX; pages 125–145. New York: McGraw-Hill Book Company, Inc., 1942.

Kenney, John F., *Mathematics of Statistics.* Part II, Chaps. VI–VII; pages 24–29, 172–186. New York: D. Van Nostrand Company, Inc., 1939.

Lindquist, E. F., *A First Course in Statistics.* Revised Edition. Chap. VIII; pages 208–211. Boston: Houghton Mifflin Company, 1942.

Lindquist, E. F., *Statistical Analysis in Educational Research.* Chaps. I, III; pages 210–218. Boston: Houghton Mifflin Company, 1940.

Mode, Elmer B., *The Elements of Statistics.* Pages 347–356. New York: Prentice-Hall, Inc., 1941.

Monroe, Walter S., Editor, *Encyclopedia of Educational Research.* Pages 399–400, 837, 998–1000, 1185. New York: The Macmillan Company, 1941.

Monroe, Walter S., and Engelhart, Max D., *The Scientific Study of Educational Problems.* Pages 103–110. New York: The Macmillan Company, 1926.

Peters, Charles C., and Van Voorhis, Walter R., *Statistical Procedures and Their Mathematical Bases.* Chaps. V, VI; pages 324–326, 419–423. New York: McGraw-Hill Book Company, Inc., 1940.

Walker, Helen M., *Elementary Statistical Methods.* Chap. XV; pages 140–142, 246–249. New York: Henry Holt & Company, Inc., 1943.

## APPENDIX

# Tables of the Normal Curve

Tables of the normal curve and its functions may be presented in various forms, but that in which other functions are tabled with respect to the standard or median deviation is most frequently useful; therefore it is followed here. More detailed tables in this form and others organized with respect to area, height, or some other functions may be found in some of the references given in Chapter I and elsewhere.

The first of the two tables presented here gives the heights, included and excluded areas, and two ratios of areas, corresponding to each value of the standard deviation or error from .0 up to 5.0, by intervals of .1; the second does the same for the median deviation or probable error from .0 up to 6.0. The values of the two deviations just referred to are in the first columns of the tables.

The second columns contain the heights, $h$, or ordinates of the normal curve at the given deviation distances. These are given in term of a maximum height, or central ordinate, of 1.00, for which the area under the curve is $\sqrt{2\pi}$ or 2.5066. Some tables of the same function present them in terms of an area of 1.00, for which the maximum height is $\frac{1}{\sqrt{2\pi}}$ or .3989. When so presented, the symbol $z$ is commonly employed for them. Just as for the maximum height, so for any other, $z = .3989h$, hence can readily be obtained from the $h$ entries.

Each entry in the third column is the fraction of the whole area under the normal curve included between an ordinate at the corresponding distance and the maximum ordinate, which, of course, is at the mean. Each entry in the fourth column is the fraction of the area beyond the same ordinate, or excluded by it. The sum of the two is always .5000.

Each fifth-column entry is obtained by dividing the corresponding fourth-column one into that in the third column. In other words, it is the ratio of the area closer to the mean than the corresponding distance in the first column to the area farther from the mean. In appropriate situations this entry represents the chances to 1 that a single case is within rather than without an ordinate at the corresponding distance.

Finally, each entry in the last column is 1 minus the entry in the fourth divided by that in the fourth. It is also equal to one more than twice the fifth-column entry. It is the ratio of the area of the larger of the two parts into which the ordinate divides the whole area under the curve to that of the smaller. Therefore it is the chances to 1 that a given case is not beyond the corresponding distance in one, but not in both, directions. In other words, it is the same as the chances arrived at by the experimental coefficient procedure.

As an example, the row of entries beginning with a $\sigma$ distance of 1.5 in the first table may be used. The height of the curve, or length of the ordinate, $1.5\sigma$ in each direction from the mean is .3247 of its height at the mean, that is, of the maximum ordinate. If $z$ rather than $h$ is desired, it is $.3989 \times .3247 = .1295$. The area included between an ordinate at $\pm 1.5\sigma$ and the mean is .4332 of the total area and that beyond $\pm 1.5\sigma$ is .0668 of the total area. If the area between $-1.5\sigma$ and $+1.5\sigma$ is desired, it can be obtained by doubling .4332, the result being .8664; if the area farther than $1.5\sigma$ from the mean in both directions is wanted, .0668 should be doubled, giving .1336. The ratio of .4332 to .0668, or of .8664 to .1336, is approximately 6.5, which means that 6.5 times as much area is within $1.5\sigma$ as is beyond it. In other words, the chances are 6.5 to 1 that any single case is within a distance of $1.5\sigma$ from the mean. Finally, $\frac{1 - .0668}{.0668}$ or $2 \times 6.5 + 1 = 14.0$, the ratio of all the area on the larger side of an ordinate at $\pm 1.5\sigma$ from the mean to all that on the smaller side. This may be defined also as the chances to 1 that a single case lies below $1.5\sigma$ above the mean or above $1.5\sigma$ below the mean rather than the opposite.

## TABLE XLIV

HEIGHTS, AREAS, AND CORRESPONDING RATIOS OF AREAS OF A NORMAL DISTRIBUTION CORRESPONDING TO STANDARD DEVIATION OR ERROR UNITS

| $\sigma$ or $\epsilon$ Distance from Mean | $h$, Height or Ordinate at Given $\sigma$ or $\epsilon$ Distance | Area Between Ordinate at Given $\sigma$ or $\epsilon$ Distance and Mean | Area Beyond Ordinate at Given $\sigma$ or $\epsilon$ Distance from Mean | Ratio of Area Between Ordinate and Mean to Area Beyond Ordinate | Ratio of Area on Larger Side of Ordinate to Area on Smaller Side |
|---|---|---|---|---|---|
| .0 | 1.0000 | .0000 | .5000 | .00 | 1.00 |
| .1 | .9950 | .0398 | .4602 | .09 | 1.17 |
| .2 | .9802 | .0793 | .4207 | .19 | 1.38 |
| .3 | .9560 | .1179 | .3821 | .31 | 1.62 |
| .4 | .9231 | .1554 | .3446 | .45 | 1.90 |
| .5 | .8825 | .1915 | .3085 | .62 | 2.24 |
| .6 | .8353 | .2257 | .2743 | .82 | 2.65 |
| .7 | .7827 | .2580 | .2420 | 1.07 | 3.13 |
| .8 | .7261 | .2881 | .2119 | 1.36 | 3.72 |
| .9 | .6670 | .3159 | .1841 | 1.72 | 4.43 |
| 1.0 | .6065 | .3413 | .1587 | 2.1 | 5.3 |
| 1.1 | .5461 | .3643 | .1357 | 2.7 | 6.4 |
| 1.2 | .4868 | .3849 | .1151 | 3.3 | 7.7 |
| 1.3 | .4296 | .4032 | .0968 | 4.2 | 9.3 |
| 1.4 | .3753 | .4192 | .0808 | 5.2 | 11.4 |
| 1.5 | .3247 | .4332 | .0668 | 6.5 | 14.0 |
| 1.6 | .2780 | .4452 | .0548 | 8.1 | 17.2 |
| 1.7 | .2357 | .4554 | .0446 | 10.2 | 21.4 |
| 1.8 | .1979 | .4641 | .0359 | 12.9 | 26.8 |
| 1.9 | .1645 | .4713 | .0287 | 16.4 | 33.8 |
| 2.0 | .1353 | .4772 | .0228 | 21. | 43. |
| 2.1 | .1103 | .4821 | .0179 | 27. | 55. |
| 2.2 | .0889 | .4861 | .0139 | 35. | 71. |
| 2.3 | .0710 | .4893 | .0107 | 46. | 92. |
| 2.4 | .0561 | .4918 | .0082 | 60. | 121. |
| 2.5 | .0439 | .4938 | .0062 | 80. | 160. |
| 2.6 | .0340 | .49534 | .00466 | 106. | 214. |
| 2.7 | .0261 | .49653 | .00347 | 143. | 287. |
| 2.8 | .0198 | .49744 | .00256 | 195. | 390. |
| 2.9 | .0149 | .49813 | .00187 | 267. | 535. |
| 3.0 | .0111 | .49865 | .00135 | 369. | 739. |
| 3.1 | .00819 | .49903 | .00097 | 520 | 1030 |
| 3.2 | .00598 | .49931 | .00069 | 730 | 1450 |
| 3.3 | .00433 | .499517 | .000483 | 1030 | 2070 |
| 3.4 | .00309 | .499663 | .000337 | 1480 | 2970 |
| 3.5 | .00219 | .499767 | .000233 | 2150 | 4300 |
| 3.6 | .00153 | .499841 | .000159 | 3100 | 6300 |
| 3.7 | .00107 | .499892 | .000108 | 4600 | 9300 |
| 3.8 | .00073 | .499928 | .000072 | 6900 | 11300 |
| 3.9 | .00050 | .4999519 | .0000481 | 10400 | 20800 |
| 4.0 | .00034 | .4999683 | .0000317 | 15800 | 31500 |
| 4.1 | .000224 | .4999793 | .0000207 | 24200 | 48300 |

TABLE XLIV—(*Concluded*)

HEIGHTS, AREAS, AND CORRESPONDING RATIOS OF AREAS OF A NORMAL DISTRIBUTION CORRESPONDING TO STANDARD DEVIATION OR ERROR UNITS

| $\sigma$ or $\epsilon$ Distance from Mean | $h$, Height or Ordinate at Given $\sigma$ or $\epsilon$ Distance | Area Between Ordinate at Given $\sigma$ or $\epsilon$ Distance and Mean | Area Beyond Ordinate at Given $\sigma$ or $\epsilon$ Distance from Mean | Ratio of Area Between Ordinate and Mean to Area Beyond Ordinate | Ratio of Area on Larger Side of Ordinate to Area on Smaller Side |
|---|---|---|---|---|---|
| 4.2 | .000148 | .4999867 | .0000133 | 37600 | 75200 |
| 4.3 | .000097 | .4999915 | .0000085 | 59000 | 118000 |
| 4.4 | .000062 | .4999946 | .0000054 | 93000 | 185000 |
| 4.5 | .000040 | .49999660 | .00000340 | 147000 | 294000 |
| 4.6 | .000025 | .49999789 | .00000211 | 237000 | 473000 |
| 4.7 | .000016 | .49999870 | .00000130 | 384000 | 769000 |
| 4.8 | .000010 | .49999921 | .00000079 | 630000 | 1260000 |
| 4.9 | .000006 | .49999952 | .00000048 | 1040000 | 2090000 |
| 5.0 | .000004 | .49999971 | .00000029 | 1740000 | 3490000 |

TABLE XLV

HEIGHTS, AREAS, AND CORRESPONDING RATIOS OF AREAS OF A NORMAL DISTRIBUTION CORRESPONDING TO MEDIAN DEVIATION OR PROBABLE ERROR UNITS

| *MdD* or *PE* Distance from Mean | *h*, Height or Ordinate at Given *MdD* or *PE* Distance | Area Between Ordinate at Given *MdD* or *PE* Distance and Mean | Area Beyond Ordinate at Given *MdD* or *PE* Distance from Mean | Ratio of Area Between Ordinate and Mean to Area Beyond O dinate | Ratio of Area on Larger Side of Ordinate to Area on Smaller Side |
|---|---|---|---|---|---|
| .0 | 1.0000 | .0000 | .5000 | .00 | 1.00 |
| .1 | .9977 | .0269 | .4731 | .06 | 1.11 |
| .2 | .9909 | .0537 | .4463 | .12 | 1.24 |
| .3 | .9797 | .0802 | .4198 | .19 | 1.38 |
| .4 | .9643 | .1063 | .3937 | .27 | 1.54 |
| .5 | .9447 | .1320 | .3680 | .36 | 1.72 |
| .6 | .9214 | .1571 | .3429 | .46 | 1.92 |
| .7 | .8945 | .1816 | .3184 | .57 | 2.14 |
| .8 | .8645 | .2053 | .2947 | .70 | 2.39 |
| .9 | .8317 | .2281 | .2719 | .84 | 2.68 |
| 1.0 | .7965 | .2500 | .2500 | 1.00 | 3.00 |
| 1.1 | .7594 | .2709 | .2291 | 1.18 | 3.37 |
| 1.2 | .7207 | .2909 | .2091 | 1.39 | 3.78 |
| 1.3 | .6808 | .3097 | .1903 | 1.63 | 4.26 |
| 1.4 | .6403 | .3275 | .1725 | 1.90 | 4.80 |
| 1.5 | .5994 | .3442 | .1558 | 2.21 | 5.42 |
| 1.6 | .5586 | .3597 | .1403 | 2.56 | 6.13 |
| 1.7 | .5182 | .3742 | .1258 | 2.98 | 6.95 |
| 1.8 | .4785 | .3876 | .1124 | 3.5 | 7.9 |
| 1.9 | .4399 | .4000 | .1000 | 4.0 | 9.0 |
| 2.0 | .4026 | .4113 | .0887 | 4.6 | 10.3 |
| 2.1 | .3667 | .4217 | .0783 | 5.4 | 11.8 |
| 2.2 | .3326 | .4311 | .0689 | 6.3 | 13.5 |
| 2.3 | .3002 | .4396 | .0604 | 7.3 | 15.6 |
| 2.4 | .2698 | .4473 | .0527 | 8.5 | 18.0 |
| 2.5 | .2413 | .4541 | .0459 | 9.9 | 20.8 |
| 2.6 | .2149 | .4603 | .0397 | 11.6 | 24.2 |
| 2.7 | .1905 | .4657 | .0343 | 13.6 | 28.2 |
| 2.8 | .1681 | .4705 | .0295 | 16.0 | 32.9 |
| 2.9 | .1476 | .4748 | .0252 | 18.8 | 38.6 |
| 3.0 | .1291 | .4785 | .0215 | 22.2 | 45.5 |
| 3.1 | .1124 | .4817 | .0183 | 26. | 54. |
| 3.2 | .0974 | .4846 | .0154 | 31. | 64. |
| 3.3 | .0840 | .4870 | .0130 | 37. | 76. |
| 3.4 | .0721 | .4891 | .0109 | 45. | 91. |
| 3.5 | .0616 | .4909 | .0091 | 54. | 109. |
| 3.6 | .0524 | .49241 | .00759 | 65. | 131. |
| 3.7 | .0444 | .49371 | .00629 | 79. | 158. |
| 3.8 | .0375 | .49481 | .00519 | 95. | 192. |
| 3.9 | .0314 | .49574 | .00426 | 116. | 234. |
| 4.0 | .0263 | .49651 | .00349 | 142. | 286. |
| 4.1 | .0219 | .49716 | .00284 | 175. | 351. |

## TABLE XLV—*Concluded*

HEIGHTS, AREAS, AND CORRESPONDING RATIOS OF AREAS OF A NORMAL DISTRIBUTION CORRESPONDING TO MEDIAN DEVIATION OR PROBABLE ERROR UNITS

| *MdD* or *PE* Distance from Mean | *h*, Height or Ordinate at Given *MdD* or *PE* Distance | Area Between Ordinate at Given *MdD* or *PE* Distance and Mean | Area Beyond Ordinate at Given *MdD* or *PE* Distance from Mean | Ratio of Area Between Ordinate and Mean to Area Beyond Ordinate | Ratio of Area on Larger Side of Ordinate to Area on Smaller Side |
|---|---|---|---|---|---|
| 4.2 | .0181 | .49769 | .00231 | 216. | 433. |
| 4.3 | .0149 | .49814 | .00186 | 267. | 535. |
| 4.4 | .0122 | .49850 | .00150 | 332. | 666. |
| 4.5 | .01000 | .49880 | .00120 | 416. | 831. |
| 4.6 | .00812 | .499041 | .000959 | 520 | 1040 |
| 4.7 | .00657 | .499238 | .000762 | 660 | 1310 |
| 4.8 | .00530 | .499397 | .000603 | 830 | 1660 |
| 4.9 | .00425 | .499525 | .000475 | 1050 | 2100 |
| 5.0 | .00339 | .499627 | .000373 | 1340 | 2680 |
| 5.1 | .00269 | .499709 | .000291 | 1720 | 3440 |
| 5.2 | .00213 | .499774 | .000226 | 2210 | 4420 |
| 5.3 | .00168 | .499825 | .000175 | 2850 | 5700 |
| 5.4 | .00132 | .499865 | .000135 | 3700 | 7400 |
| 5.5 | .00103 | .499896 | .000104 | 4800 | 9600 |
| 5.6 | .000798 | .4999207 | .0000793 | 6300 | 12600 |
| 5.7 | .000617 | .4999396 | .0000604 | 8300 | 16600 |
| 5.8 | .000478 | .4999542 | .0000458 | 10900 | 21900 |
| 5.9 | .000364 | .4999655 | .0000345 | 14500 | 29000 |
| 6.0 | .000278 | .4999741 | .0000259 | 19300 | 38600 |

# Index